Anteckningar - MT5018

Sebastijan Babic

26 May 2025

Sammanfattning

Syftet med detta dokument är att ge en koncis sammanfattning av kursen MT5018 - Sannolikhetsteori II VT25. Viss notation kan vara inkonsekvent.

Innehåll

1. Repetition

Innehåll: Kursboken: Avsnitt "Introduction"

För att ha några allmänna regler inkluderar vi sannolikhetsrummet. Notation:

$$(\Omega, \mathcal{F}, \mathbb{P})$$

Där Ω är utfallsrummet, dvs. samlingen av utfall ω, $\mathcal{F}$ är samlingen av händelser, dvs. $\mathcal{F} = \{A, B, C...\}$ och $\mathbb{P}$ är sannolikhetsfunktionen. Sannolikhetsfunktionen måste uppfylla Kolmogorovs axiom:

1. För varje $A \in \mathcal{F}$ finns det $\mathbb{P}(A)$ sådan att $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(\Omega) = 1$
3. Låt $\{A_n, n \geq 1\}$ vara en samling av parvis disjunkta händelser, och A vara deras union. Då gäller

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

Oberoende: A, B är oberoende om och endast om

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Longleftrightarrow \mathbb{P}(A \mid B) = P(A)$$

Om detta är sant, då är A, B parvis oberoende.

Notera att detta inte är samma som disjunkta. Disjunkta om

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$

gäller.

Example: Tag $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 2\}$ och $B = \{1, 3, 4\}$, då är dem oberoende eftersom

$$\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

vilket stämmer med att $\mathbb{P}(A \cap B) = \mathbb{P}(\omega : \{1\}) = \frac{1}{6}$.

Men dem är inte disjunkta eftersom 1 är i både A och B.

Betingad sannolikhet: Om $\mathbb{P}(B) > 0$, då är den betingade sannolikheten

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

Observera att om händelserna är oberoende får vi $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Lagen om total sannolikhet:

Låt $\{H_k, 1 \leq k \leq n\}$ vara en partition av Ω, dvs. disjunkta mängder och $\cup_{k=1}^n H_k = \Omega$, då gäller

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(A \mid H_k)\mathbb{P}(H_k)$$

Som en konsekvens får vi

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(A \mid H_i)\mathbb{P}(H_i)}{\sum_{k=1}^n \mathbb{P}(A \mid H_k)\mathbb{P}(H_k)}$$

i Bayes formel. Detta är lagen om total sannolikhet.

Notera att vi kan skriva om lagen så att om A en händelse och Y en indikator variabel, dvs $Y = 1$ om händelsen sker, 0 annars. Då är $\mathbb{E}[Y] = \mathbb{P}(A) = 1 \cdot p_{Y(1)}$, då fås

$$\mathbb{P}(A) = \mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(\mathbb{P}(A \mid X)) = \sum_x \mathbb{P}(A \mid X = x) p_X(x)$$

Kontinuerligt skriver vi då

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x)$$

1.1. Stokastiska variabler

En stokastisk variabel är en funktion

$$X : \Omega \to \mathbb{R}$$

Vi kan beskriva dess sannolikhetsmassafunktion via en fördelningsfunktion

$$F_{X(x)} = \mathbb{P}(X \leq x) \quad -\infty < x < \infty$$

I denna kurs kommer vi endast att betrakta diskreta fördelningar och kontinuerliga fördelningar. De har båda sina egna fördelningsfunktioner:

$$\begin{cases} \mathbb{P}(X = x) = \begin{cases} p_X(x) \\ f_X(x) \end{cases} \\ \\ F_{X(x)} = \begin{cases} \sum_{y \leq x} p_X(y) \\ \int_{-\infty}^{x} f_X(y) \end{cases} \end{cases}$$

För $-\infty < x < \infty$.

Väntevärde: Väntevärdet kan beskrivas via

$$\mathbb{E}[X] = \begin{cases} \sum_{x_k} x_k p_X(x_k) \\ \int_{-\infty}^{\infty} x f_X(x) dx \end{cases}$$

(givet absolut konvergens). Kom ihåg de stora talens lag, om X_i är oberoende och identiskt fördelade, då gäller

$$\frac{\sum X_i}{n} \to \mathbb{E}[X_i]$$

Median: För data som kan ha mycket stora avvikelser, dvs. extrema värden vid några få punkter, kan vi använda medianen

$$\begin{cases} \mathbb{P}(X \geq m) \geq \frac{1}{2} \\ \mathbb{P}(X \leq m) \leq \frac{1}{2} \end{cases}$$

varians:

$$V[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Detta är den vanliga formeln för varians.

Moment:

$$\mathbb{E}[X^n]$$

här är n den n:te momentet.

Simultan fördelningsfunktion: För varje stokastisk vektor $(\boldsymbol{X}) = (X_1, ..., X_n)$ kan vi beskriva den simultana fördelningsfunktionen som

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq x) = \begin{cases} p_{\boldsymbol{X}}(\boldsymbol{x}) \\ f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^n F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial \boldsymbol{x}} \end{cases}$$

Simultan oberoende: Om simultan oberoende gäller, då har vi

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1}(x_1)...F_{X_n}(x_n)$$

Om simultan oberoende är uppfyllt, då gäller

$$\mathbb{E}[\boldsymbol{X}] = \mathbb{E}[X_1]\mathbb{E}[X_2]...\mathbb{E}[X_n]$$

Kovarians: För att mäta kovariansen använder vi

$$\operatorname{cov}[X] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Korrelationskoefficienten definieras som:

$$\rho_X = \frac{\operatorname{Cov}(X, Y)}{\sqrt{V[X]V[Y]}}$$

Observera att $\rho_{X,Y} \in [-1, 1]$, om 0 säger vi att variablerna är okorrelerade.

2. Föreläsning - Flerdimensionella stokastiska variabler

Innehåll: Kursboken: Kapitel 1.1-1.2.

Definition: En n-dimensionell stokastisk variabel eller vektor $\boldsymbol{X}$ är en (mätbar) funktion från sannolikhetsrummet till $\mathbb{R}^n$, dvs.

$$\boldsymbol{X} : \Omega \to \mathbb{R}^n$$

där $\boldsymbol{X} = (X_1, ..., X_n)^T$. varje X_i är en komponent.

Simultan fördelningsfunktion: Dess simultana fördelningsfunktion är

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x})$$

komponentvis. Så vi får

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \leq x) = \begin{cases} p_{\boldsymbol{X}}(\boldsymbol{x}) \\ f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\partial^n F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial \boldsymbol{x}} \end{cases}$$

Marginalfördelning: För det diskreta fallet får vi

$$p_X(x) = \mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y)$$

respektive för det kontinuerliga fallet får vi

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

Marginalfördelningsfunktionen blir

$$F_{X(x)} = \int_{-\infty}^{x} f_X(u) du = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(u, y) dy \right) du$$

Transformationssatsen: Transformationssatsen ger ett sätt att hitta täthetsfunktionen (pdf) för en ny stokastisk vektor som är en funktion av en annan stokastisk vektor med känd pdf.

Låt $\boldsymbol{X}$ vara en n-dimensionell kontinuerlig stokastisk vektor med täthet $f_{\boldsymbol{X}}(\boldsymbol{x})$ koncentrerad på en mängd $S \subseteq \mathbb{R}^n$. Låt $\boldsymbol{g} = (g_1, g_2, ..., g_n)$ vara en bijektiv (injektiv och surjektiv) funktion från S till någon mängd $T \subseteq \mathbb{R}^n$. Betrakta den n-dimensionella stokastiska vektorn $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$, vilket betyder $Y_i = g_{i(X_1, ..., X_n)}$ för $i = 1, ..., n$. Antag att $\boldsymbol{g}$ och dess invers $\boldsymbol{h} = \boldsymbol{g}^{-1}$ båda är kontinuerligt deriverbara. Tätheten för $\boldsymbol{Y}$ ges av:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \begin{cases} f_{\boldsymbol{X}}\big(h_1(\boldsymbol{y}), h_2(\boldsymbol{y}), ..., h_{n(\boldsymbol{y})}\big) \cdot |J| & \text{för } \ \boldsymbol{y} \in T \\ 0 & \text{annars} \end{cases}$$

där $\boldsymbol{h} = (h_1, ..., h_n)$ är den unika inversen till $\boldsymbol{g}$, och J är Jacobideterminanten av inverstransformationen $\boldsymbol{h}$:

$$J = \det \begin{pmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \cdots & \frac{\partial h_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \cdots & \frac{\partial h_n}{\partial y_n} \end{pmatrix}$$

ionen av en täthetsfunktion måste integranden vara tätheten $f_{\boldsymbol{Y}}(\boldsymbol{y})$.

2.0.1. Enkla exempel

Summa och skillnad av normalfördelade variabler: Låt X och Y vara oberoende $N(0, 1)$ stokastiska variabler. Hitta den simultana fördelningen av $U = X + Y$ och $V = X - Y$.

Inverstransformationen är $X = \frac{U+V}{2}$ och $Y = \frac{U-V}{2}$. Jacobianen är $J = -\frac{1}{2}$, så $|J| = \frac{1}{2}$. Den simultana tätheten för X och Y är $f_{X,Y}(x,y) = f_X(x) f_Y(y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right)$. Genom att tillämpa satsen:

$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \frac{1}{2}$$

$$= \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(\left(\frac{u+v}{2}\right)^2 + \left(\frac{u-v}{2}\right)^2\right)\right) \cdot \frac{1}{2}$$

$$= \frac{1}{2\pi} \exp\left(-\frac{1}{2}\frac{u^2+v^2}{2}\right) \cdot \frac{1}{2}$$

$$= \left(\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{u^2}{2\cdot 2}}\right)\left(\frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{v^2}{2\cdot 2}}\right)$$

Detta visar att U och V är oberoende $N(0,2)$ stokastiska variabler.

Kvot och summa av exponentialfördelade variabler: Låt X och Y vara oberoende $\mathrm{Exp}(1)$ stokastiska variabler. Hitta den simultana fördelningen av $U = \frac{X}{X+Y}$ och $V = X + Y$.

Inverstransformationen är $X = UV$ och $Y = V - UV = V(1-u)$. Jacobianen är $J = v$, så $|J| = v$ (eftersom $V = X + Y > 0$). Den simultana tätheten för X och Y är $f_{X,Y}(x,y) = e^{-x}e^{-y} = e^{-(x+y)}$ för $x, y > 0$. Genom att tillämpa satsen för $u \in (0,1)$ och $v > 0$:

$$f_{U,V}(u,v) = f_{X,Y}(uv, v(1-u)) \cdot v = e^{-(uv+v(1-u))} \cdot v = e^{-v} \cdot v$$

Denna täthet faktoriseras som $1 \cdot (ve^{-v})$, vilket visar att $U \in U(0,1)$ och $V \in \Gamma(2,1)$ är oberoende.

2.0.2. Many-To-One-transformationer (kan ignoreras)

Many-To-One-transformationer: Om funktionen $\boldsymbol{g}$ inte är injektiv behöver satsen modifieras. Anta att domänen S kan partitioneras i m disjunkta delmängder $S_1, \dots, S_m$ sådana att $\boldsymbol{g}$ begränsad till varje S_k är en bijektion till T och uppfyller satsens villkor. Låt $\boldsymbol{h}_k$ vara inversen till $\boldsymbol{g}$ begränsad till S_k, och låt J_k vara motsvarande Jacobian. Då är tätheten för $\boldsymbol{Y}$:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{k=1}^{m} f_{\boldsymbol{X}}\left(\boldsymbol{h}_{k(\boldsymbol{y})}\right) \cdot |J_k|, \quad \text{för} \quad \boldsymbol{y} \in T$$

Kvadrattransformation: För $y = x^2$ är funktionen 2-till-1 på $\mathbb{R} \setminus \{0\}$. Vi har $S_1 = (0, \infty)$ med invers $h_1(y) = \sqrt{y}$ och $S_2 = (-\infty, 0)$ med invers $h_2(y) = -\sqrt{y}$. Jacobimagnituden är för båda $|\frac{d}{dy}(\pm\sqrt{y})| = \frac{1}{2\sqrt{y}}$. Formeln ger:

$$f_{Y(y)} = f_{X(\sqrt{y})} \ |\frac{1}{2\sqrt{y}}| + f_{X(-\sqrt{y})} \ |\frac{1}{2\sqrt{y}}| = \frac{f_{X(\sqrt{y})} + f_{X(-\sqrt{y})}}{2\sqrt{y}}$$

3. Föreläsning - Betingade fördelningar och väntevärde

Innehåll: Kursboken: Kapitel 2.1-2.2.

Betingade fördelningen: Låt (X, Y) vara en stokastisk vektor så att $\mathbb{P}(X = x) > 0$. Den betingade sannolikhets-funktionen för $Y \mid X = x$ är

$$p_{X|Y}(y) = \mathbb{P}(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_{X(x)}}$$

Analogt för det kontinuerliga fallet.

Betingad väntevärde: Definierar den betingade väntevärdet som

$$\mathbb{E}[Y \mid X = x] = \left\{\sum_y y p_{Y|X=x}, \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy\right.$$

Betingad varians: Skriver den betingade variansen som (med härledning som inte tas med här)

$$\mathrm{Var}(Y \mid X = x) = \mathbb{E}(\mathbb{V}(Y \mid X = x)) + \mathrm{Var}(\mathbb{E}(Y \mid X = x))$$

Vi skriver $\mathbb{E}[Y \mid X = x] = h(X)$ respektive $\mathbb{E}[Y \mid X] = h(X)$.

Egenskaper hos den betingade väntevärde:

1. $\mathbb{E}[g(X)Y \mid X] = g(X)\mathbb{E}[Y \mid X]$
2. $\mathbb{E}[Y \mid X] = \mathbb{E}[Y]$ (endast om oberoende)

Lagen om itererade väntevärde:

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y]$$

om $\mathbb{E}[|Y|] < \infty$

4. Föreläsning - Mer om betingning

Innehåll: Kursboken: Kapitel 2.3-2.5. Ej avsnitt 3 från (3.6) till slutet, ej Sats 5.3.

4.1. Fördelningar med slumpmässiga parametrar

Vi måste först förstå gammafunktionen lite. Inte så mycket men bara en grund.

Definition: Vet att om $Z \sim \Gamma(p, a)$ där p är formparametern och a är skalningsparametern så har vi

$$f_Z(z) = cz^{p-1}e^{-\frac{z}{a}} \quad z > 0$$

Så har något konstant c där

$$c = \frac{1}{\Gamma(p)a^p}$$

där nämnarens Γ fås via

$$\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx$$

som är en konvergent integral för alla $p > 0$. Kan även skriva funktionen som

$$\Gamma(n) = (n-1)!$$

Vi använder ibland detta för att känna igen en integral som en gammaintegral och därmed få integranden att likna en gammatäthet.

Example: Låt N vara antalet partiklar $\sim \mathrm{Po}(\lambda)$. Låt X vara antalet detekerade partiklar/timme. Låt $X \mid N = n \sim \mathrm{Bin}(n, p)$

- Vad är då fördelningen för X?

Vi har

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{n=0}^{\infty} \mathbb{P}(X = k \mid N = n)p_N(n) \\
&= \sum_{n=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\
&= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n-k)!} (1-p)^{n-k}
\end{aligned}$$

som genom Taylor expansion av e ger vid substition $j = n - k$ sannolikheten

$$\frac{(\lambda p)^k}{k!} e^{-\lambda p}$$

och därmed fås slutsatsen $X \sim \mathrm{Po}(\lambda p)$.

4.2. Betingning

Vi har den välkända Bayes formel.

Theorem: Om vi har en diskret/kontinuerlig slumpvektor (X, Y) så är deras betingade sannolikhetsfunktioner

$$\begin{cases} p_{Y|X}(y) = \frac{p_{X|Y}(x,y)}{p_{X(x)}} \\ f_{Y|X} = \frac{f_{X|Y}(x,y)}{f_{X(x)}} \end{cases}$$

Om vi har en blandat form där X diskret och Y kontinuerlig så fås följande.

Theorem:

$$F_{Y|X=k}(y) = \frac{\mathbb{P}(Y \leq y, X = k)}{\mathbb{P}(X = k)}$$

eller alternativt på ett mer användbart sätt:

$$f_{Y|X=k}(y) = \frac{\mathbb{P}(X = k \mid Y = y) f_{Y(y)}}{\mathbb{P}(X = k)}$$

Vi behöver ibland även Beta-fördelningen som vi ska bara känna igen ibland.

Definition: $Z \sim \beta(r, s)$ om

$$f_Z(z) = cz^{r-1}(1-z)^{s-1}, \quad z \in (0, 1)$$

We say

$$\int_0^1 f_Z(z) dz = 1, c = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}$$

5. Föreläsning - Sannolikhetsgenererande funktioner

5.1. Lite om prediktion (kan ignoreras)

Definition: En prediktor för Y baserade på X är en funktion $d(X)$

Definition: Kvadratisk prediktionsfel

$$\mathbb{E}\left[(Y - d[X])^2\right]$$

vi definerar

$$h(X) = \mathbb{E}[Y \mid X = x]$$

Theorem: $h(X)$ är den bästa prediktorn. Minimerar kvadratiska prediktionsfelet.

5.2. Transformer (här sannolikhetsgenererande funktioner)

Noterar först att det man transformerar är fördelningar, beteckande $L(X)$ och inte X självt. Vi skriver

$$X \stackrel{d}{=} Y$$

om samma förldening. Alternativt

$$L(X) = L(Y)$$

Notera även att om samma fördelning, dvs. $X \stackrel{d}{=} Y$ betyder det inte $X(w)Y(w)$.

Example: Singlar slant. Låt

$$X = \begin{cases} 1 \text{ om krona} \\ 0 \text{ annars} \end{cases}, \quad Y = \begin{cases} 1 \text{ om klöver} \\ 0 \text{ annars} \end{cases}$$

Dessa två händelser kommer aldrig ske samtidigt även om dem har samma fördelning, dvs. $X, Y \sim \text{Be}\left(\frac{1}{2}\right)$.

Definition: Om X diskret på $\{0, 1, 2, ...\}$ så säger vi att dess sannolikhetsgenererande funktion är

$$g_{X(t)} := \mathbb{E}\left[t^X\right] = \sum_{k=0}^{\infty} t^k \mathbb{P}(X = k)$$

Theorem: Vi kan få sannolikheten för att händelsen X antar ett visst värde genom

$$\mathbb{P}(X = n) = \frac{g_X^{(n)}(0)}{n!}$$

där n är den n:te derivata. Notera att vi antar godtyckligt deriverbar.

Corollary: Följd av ovanstående sats är att om två variabler X, Y har samma sannolikhetsgenererande funktion så har dem även samma fördelning. Dvs.

$$g_X = g_Y \Leftrightarrow L(X) = L(y)$$

Vid derivering får vi $f(X) = X(X-1)(X_2)...(X-n+1)$ så får vi följande sats

Theorem: Om $\mathbb{E}[X^n] < \infty$ så

$$g_X^{(n)}(1) = \mathbb{E}[f(X)]$$

Corollary:

1. $\mathbb{E}[X] = g_X'(1)$

2. $\text{Var}[X] = g''_X(1) + g'_X(1) - (g'_X(1))^2$

5.3. Faltning

Antag slumpvariabler $X_1, X_2, ..., X_n$, X_i oberoende för alla i. Skriver $S = X_1 + ... + X_n$. Sannolikhetsgenererande funktionen för S_n är då följande.

Theorem:

$$g_{S_n}(t) = \mathbb{E}\left[t^{S_n}\right] = \prod_{i=1}^{n} g_{X_i}(t)$$

Corollary: Om följden $\{X_i\}$ oberoende för alla $i \leq n$ och även likafördelade så gäller det att

$$g_{S_n}(t) = \left(g_{X(t)}\right)^n$$

6. Föreläsning - Momentgenererande (och karakteriska men ignoreras) funktioner

Innehåll : Kursboken: Kapitel 3.1-3.2.

6.1. Momentgenererande funktioner

Problemet med sannolikhetsgenererande funktionen är att den är begränad på positiva heltalen $\{1, 2, ...\}$.

Noterar först att för positiva heltal så gäller det att

$$\psi_X(t) = g_X(e^t) \quad |t| < h$$

givet att sannolikhetsgenererande funktionen existerar för $|t| < h$.

Theorem: Om X, Y slumpvariabler och om $\exists h > 0$ så att

$$\psi_X(t) = \psi_Y(t) \quad |t| < h, h > 0$$

så gäller det att

$$X \overset{d}{=} Y$$

6.1.1. Faltning

Theorem: Om följden $\{X_i\}$ är oberoende (och likafördelade?) så gäller det att för $S_n = X_1 + ... + X_n$ att

$$\psi_{S_n}(t) = \prod_{k=1}^{\infty} \psi_{X(i)}(t) \quad |t| < h$$

Corollary: Om X_i likafördelade ch oberoende för alla i så gäller det att

$$\psi_{S_n}(t) = \left(\psi_{X_i}(t)\right)^n \quad |t| < h$$

Theorem: Om det existerar en momentgenererande funktion $\psi_X(t)$ så gäller det att

1. Alla moment existerar, dvs. $\mathbb{E}[|X|^r] < \infty \quad \forall r > 0$
2. $\mathbb{E}[X^n] = \psi_X^{(n)}(0) \quad n = 1, 2, ...$

Noterar speciellt att första momentet är väntevärdet, andra momentet är variansen.

Theorem: Om X slumpvariabel och $a, b \in \mathbb{R}$ så gäller det att

$$\psi_{aX+b}(t) = e^{tb}\psi_X(at)$$

Definition: Om vi har en slumpvektor $\boldsymbol{X} = (X_1, ..., X_n)^T$ så gäller det att

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[e^{\boldsymbol{t}\boldsymbol{X}}]$$

givet att $\boldsymbol{h} > 0$ så att väntevärdet existerar för $|t_k| < h_k$ där $k = 1, 2, ...$

Example: Om $X \sim U(a, b)$, då är

$$\psi_X(t) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

omm. $t \neq 0$.

7. Föreläsning - Mer om transformer

Innehåll: Kursboken: Kapitel 3.5-3.7.

7.1. Fortsättning faltning

Theorem: Om $N, X_1, \ldots$ oberoende slumpvariabler som antar värde i mängden $\{0, 1, 2, \ldots\}$ där $\{X_i\}$ iid. så gäller det att

$$g_{S_N}(t) = g_N(g_X(t))$$

Theorem: Om $\{X_i\}$ iid så att det existerar $\psi_{X_i}(t) \quad \forall |t| < h$ och låt $N \in \{0, 1, 2, \ldots\}$ oberoende av $\{X_i\}$, då gäller det att

$$\psi_{S_N}(t) = g_N(\psi_X(t))$$

Theorem: Om $N, X_1, \ldots$ oberoende slumpvariabler på $\{0, 1, 2, \ldots\}$ där $\{X_i\}$ iid. så gäller det att

1. Om $\mathbb{E}[N] < \infty$ och $\mathbb{E}[X] < \infty$ så gäller det att

$$\mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[X]$$

2. Om $\operatorname{Var}[X], \operatorname{Var}[N] < \infty$ så gäller det att

$$\operatorname{Var}[S_N] = \mathbb{E}[N] \operatorname{Var}[X] + \mathbb{E}[X]^2 \operatorname{Var}[N]$$

$$\operatorname{Var}[S_n] = \operatorname{Var}\left(\sum_{i=1}^{N} X_i \mid N\right) = E\left[\operatorname{Var}\left[\sum_{i=1}^{N} X_i \mid N = n\right]\right] + \operatorname{Var}\left[E\left[\sum_{i=1}^{N} X_i \mid N = n\right]\right] = E[N] \operatorname{Var}[X] + \operatorname{Var}[E[X]N]] =$$
$$= E[N] \operatorname{Var}[X] + E[X]^2 \operatorname{Var}[N]$$

7.2. Förgreningsprocesser

Definition: Om vi låter generation n vara avkomma till individ i generation $n-1$ och desstom att individen ger upphov till avkomma:

1. Enligt samma fördelning $\stackrel{d}{=} Y$
2. Oberoende

Vi vill ha $X(n)$=antalet individier generation n med $X(0) = 1$ och om vi låter Y_i^j =antalet barn till individ i generation j så är

$$X(n) = \sum_{i=1}^{X(n-1)} Y_i^{n-1}$$

där $\mathbb{E}[Y] = m$

Theorem:

$$\mathbb{E}[X(n)] = m^n$$

Definition: Låter $\eta = \mathbb{P}(\text{dör ut}) = \mathbb{P}(X(n) = 0 \text{ något } n)$

Theorem:

1. η är den minsta icke-negativa roten till ekvationen

$$t = g_Y(t)$$

1. Om $m \leq 1$ så innebär det att $\eta = 1$, men om $m > 1$ så innebär det att $\eta < 1$.

8. Föreläsning - Kovariansmatriser samt den multivariata normalfördelningen

Innehåll: Kursboken: Kapitel 5.1-5.3.

8.1. Lite linjär algebra

Ortogonoal: En matris $\mathbb{C}$ är ortogonal omm

$$\mathbb{C}\mathbb{C}^T = \mathbb{I}$$

Corollary: Om A symmetrisk så existerar det ett ortogonal matris så att

$$A = \mathbb{C}D\mathbb{C}^T, \quad D = \operatorname{diag}(\lambda_1, ..., \lambda_N)$$

Här är λ egenvärden till A. Det gäller att

$$\det(A) = \det(D) = \prod_{i=1}^{n} \lambda_i$$

När vi säger att en matris är ortogonal betyder det i ord att raderna och kolumnerna, betraktade som vektorer är ortonormala, det vill säga har dem längd 1 och är ortogonala alltså skalär produkten mellan dem är 0.

Theorem: Om $\mathbb{A}$ har en invers så gäller det att

$$\left(\mathbb{A}^{-1}\right)^{\frac{1}{2}} = \left(\mathbb{A}^{\frac{1}{2}}\right)^{-1}$$

8.2. Multivariata normalfördelningen

Betrakta den flerdimensionella slumpvariabeln

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix}$$

Väntevärdesvektor: Definierar väntevärdesvektorn som

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{pmatrix}$$

där $\mu_i = \mathbb{E}(X_i)$

Kovariansmatrisen: Definierar kovariansmatrisen som

$$\Lambda = \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right]$$

dvs har vi formen

$$\Lambda = \mathbb{E}\left[\begin{pmatrix} \vdots \\ X_i - \mu_i \\ \vdots \end{pmatrix}, \begin{pmatrix} \dots & X_j - \mu_j & \dots \end{pmatrix}\right]$$

Vi skriver då

$$\Lambda = \operatorname{Cov}(\boldsymbol{X})$$

Note: Värt att notera att Λ är symmetrisk matris och är även alltid icke negativ definit,dvs. gäller det att

$$\boldsymbol{a}^T \Lambda \boldsymbol{a} = \operatorname{Var}(Y) \geq 0$$

eller mer formellt utifrån linjär algebra gäller det att en $n \times n$-matris M är icke negativt definit om och endast om

$$\boldsymbol{a}T\Lambda\boldsymbol{a} \geq 0 \quad \forall \text{ n-vektorer där } \quad \boldsymbol{a} \neq$$

Då kan vi även visa att om icke-negativ definit så är även alla egenvärden ≥ 0.

Note:
1. $\det(\Lambda) = 0 \Leftrightarrow \boldsymbol{X}$ singulär
2. $\det(\Lambda) > 0 \Leftrightarrow \boldsymbol{X}$ äkta $n - \dim$

Egenskaper hos väntevärdesvektorn och kovariansmatrisen: Låt oss ha en väntevärdesvektor och en kovarianmatris. Låt oss även införa en $m \times n$ matris B och en m-vektor $\boldsymbol{b}$. Då har vi följande egenskaper
1. $\mathbb{E}(\boldsymbol{Y}) = B\boldsymbol{\mu} + \boldsymbol{b}$
2. $\operatorname{Cov}(\boldsymbol{Y}) = B\Lambda B^T$

Specialfall av tidigare sats: Om $m = 1$ och om X oberoende så implicerar det att

$$\operatorname{Var}(Y) = \sum_i a_i^2 \operatorname{Var}(X_i)$$

Eller helt enkelt sagt reduceras det till det välkända faktum att

$$\mathbb{E}(Y) = a\mathbb{E}(X) + b$$

och

$$\operatorname{Var}(Y) = a^2 \operatorname{Var}(X)$$

om $Y = aX + b$.

Vi skriver nu den första definitionen av den multivariata normalfördelningen. Den kan skrivas om med hjälp av karakteristiska funktionen och täthetsfunktionen. Vi definierar nu med hjälp av matris-egenskaper.

8.3. Första definitionen

Första definitionen av normal via multivariata normalfördelningen: Slumpvektorn $\boldsymbol{X}$ av storlek n är normal omm det för varje n-vektor $\boldsymbol{a}$ gäller att $\boldsymbol{a}^T\boldsymbol{X}$ är normal.

Vi skriver den multivariata normalfördelningen som

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Lambda)$$

Följder av ovanstående definition:
1. varje komponent av $\boldsymbol{X}$ är normal.
2. $X_1, \ldots, X_n$ är normal.
3. varje marginell fördelning är normal.
4. Om $\boldsymbol{X}$ har oberoende normala komponenter så är $\boldsymbol{X}$ normal.

Definition: Vi kan hitta $\boldsymbol{X}$ via

$$\boldsymbol{X} = \Lambda^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu}$$

där $Z_i \sim N(0, 1)$ iid.

Theorem: Antag att $\boldsymbol{X}$ har en multivariat normalfördelning och låt $\boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$. Då gäller det att

$$\boldsymbol{Y} \sim N(B\boldsymbol{\mu} + \boldsymbol{b}, B\Lambda B^T)$$

9. Föreläsning - Alternativa definitioner av den multivariata normalfördelningen

Innehåll: Kursboken: Kapitel 5.4-5.6.

9.1. MGF för multivariata N-fördelningen

Om $X \sim N(\mu, \sigma^2)$ så är

$$\psi_{X(s)} = e^{\mu s + \frac{s^2\sigma^2}{2}}$$

MGF för slumpvektorn $\boldsymbol{X}$ blir då

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) := \mathbb{E}\left(e^{\boldsymbol{t}^T\boldsymbol{X}}\right)$$

i exponenten står det alltså

$$\boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{X} = (t_1, ..., t_n) \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = t_1X_1 + ... + t_nX_n$$

Lite tydligare då så har vi

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = \psi_{\boldsymbol{X}}(t_1, ..., t_n) = \mathbb{E}(e^{t_1X_1+...+t_nX_n})$$

Antag nu att $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Lambda)$. Om vi fixerar t, då fås

$$Y = \boldsymbol{t}^T\boldsymbol{X} \sim N(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{\Lambda t}))$$

Detta är en 1-dim momentgenererande funktion. Skriver då alltså

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = E(e^{\boldsymbol{Y}}) = \psi_{\boldsymbol{Y}}(1)$$

Då fås enligt tidigare

$$= e^{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T\Lambda\boldsymbol{t}}{2}}$$

Theorem: Om $\boldsymbol{X}$ har multivariat normalfördelning med paramternarna $\boldsymbol{\mu}$ och Λ, då gäller det att

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = e^{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{\boldsymbol{t}^T\Lambda\boldsymbol{t}}{2}}$$

Det är ett alternativt sätt att definiera multivariata normalfördelningen.

Viktigaste delen, annat kan läsas "kursivt"

9.1.1. Tätheten

Om $\boldsymbol{Z} \sim N(\boldsymbol{0}, \mathbb{I})$, och om Z_i iid. så har vi

$$\prod_{i=1}^{n} f_{Z_i}(z_i) = \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^{n} e^{-\frac{z_i^2}{2}}$$

skriver om lite och får

$$\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n} z_i^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}}$$

Godtycklig $\boldsymbol{\mu}$ och Λ så att $\det(\Lambda) > 0$. Använder transformationssatsen från kapitel 2.

9.2. Ytterligare linjär algebra

Λ är positivt definit, dvs. $\det(\Lambda) > 0$ och symmetrisk. Kan skriva

$$\Lambda = \mathbb{C}D\mathbb{C}^T, \quad \Lambda^{\frac{1}{2}} = \mathbb{C}D^{\frac{1}{2}}\mathbb{C}^T$$

även $\Lambda^{\frac{1}{2}}$ är positivt definit och symmetrisk.

Eftersom Λ positivt definit så är $\det(\Lambda) > 0$ per definition av positivt defint. För en matris med denna determinant-egenskapen så betyder det att det existerar inversen Λ^{-1}. *Noterar att Λ^{-1} är positivt definit och symmetrisk.*

> Det innebär att det existerar en invers till roten, dvs.

$$\exists\left(\Lambda^{\frac{1}{2}}\right)^{-1} = (\Lambda^{-1})^{\frac{1}{2}}$$

Låt oss skriva $\Lambda^{-\frac{1}{2}}$. Denna matris är positiv definit och symmetrisk, förkortas SPD.

9.2.1. Tätheten fortsättning

Tillbaka nu till transformationssatsen. Inför $\boldsymbol{X} = g(\boldsymbol{Z})$, låt oss uttrycka då $\boldsymbol{Z} = h(\boldsymbol{X})$. Dvs. vi inverterar. Satsen säger då

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{\boldsymbol{Z}}(h(\boldsymbol{x})) \cdot |\boldsymbol{J}|$$

här är $\boldsymbol{J}$, Jacobianen följande:

$$\boldsymbol{J} = \det\left(\frac{\partial z_i}{\partial x_i}\right) = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial z_n}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_n} \end{pmatrix}$$

Egenskapen $\boldsymbol{X} = \Lambda^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu}$ ger att

$$\boldsymbol{Z} = \Lambda^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$$

Då blir Jacobianen

$$\boldsymbol{J} = \det\left(\Lambda^{-\frac{1}{2}}\right)$$

(detta är en övning i boken. Matrisen av derivator av $\boldsymbol{Z} = A\boldsymbol{X}$ är precis A), Räkneregler för determinanten ger då

$$\boldsymbol{J} = (\det(\Lambda))^{-\frac{1}{2}}$$

Nu återstår bara följande:

$$\begin{aligned} \boldsymbol{Z}^T\boldsymbol{Z} &= \left(\Lambda^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu})\right)^T \cdot \Lambda^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu}) \\ &= (\boldsymbol{x} - \boldsymbol{\mu})^T \left(\Lambda^{-\frac{1}{2}}\right)^T \Lambda^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu}) \end{aligned}$$

ty symmetrisk så fås bara

$$= (\boldsymbol{x} - \boldsymbol{\mu})^T \left(\Lambda^{-\frac{1}{2}}\right) \Lambda^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu})$$

som då ger till slut

$$= (\boldsymbol{x} - \boldsymbol{\mu})^T \Lambda^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$$

bakar ihop alltihopa och får då tätheten.

Tätheten för en slumpvektor med multivariat normalfördelning: För $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Lambda)$ med $\det(\Lambda) > 0$ gäller det att

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \underbrace{\frac{1}{\sqrt{\det(\Lambda)}}}_{\text{Jacobianen}} e^{-\frac{1}{2}((\boldsymbol{x}-\boldsymbol{\mu})^T\Lambda^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))}$$

9.2.2. Två dimensionella fallet

Har

$$\Lambda = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

och kom ihåg

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_1\sigma_2}$$

Så har

$$\det(\Lambda) = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1 - \rho^2) > 0 \quad \text{om } |\rho| < 1$$

Behöver även inversen. Så

$$\Lambda^{-1} = \frac{1}{\sigma_1^2\sigma_2^2\rho^2} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_2\sigma_2 & \sigma_1^2 \end{pmatrix}$$

Utveckling av $(\boldsymbol{x} - \boldsymbol{\mu}^T)\Lambda^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ ger (i 2-dim)

$$\frac{1}{1 - \rho^2}\left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right]$$

det ger att

$$f_{X_1,X_2}(x_1,x_2)=\frac{1}{2}\pi\cdot\frac{1}{\sigma_1\sigma_2(1-\rho^2)}e^{\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}$$

Noterar att $\boldsymbol{\mu}$=läge/position. $\boldsymbol{\sigma}$=skala. Om vi låter $\boldsymbol{\mu}=0,\boldsymbol{\sigma}=1$, detta ger

$$f(x_1,x_2)=c\cdot e^{-\frac{1}{2(1-\rho^2)}(x_1^2-2\rho x_1x_2+x_2^2)}$$

(se bilden som ges ut).

9.3. Betingning av multivariata normalfördelningen

Låt oss ha $\boldsymbol{X}\sim N(\boldsymbol{\mu},\Lambda)$ med $\det(\Lambda)>0$.

- Först $n=2$

$$f_{X_2|X_1=x_1}(x_2)=\frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}=\frac{1}{1-\rho^2}\frac{\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}{\frac{1}{2\pi\sigma_1^2}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}})$$

Detta förenklas vidare till

$$...\propto ce^{(-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(x_2-\mu_2-\rho_1\frac{\sigma_1}{\sigma_2}(x_1-x_2)^2\right)}$$

Dvs. $N\left(\mu_2+\rho\frac{\sigma_1}{\sigma_2}(x_1-\mu_1),\sigma_2^2(1-\rho^2)\right)$. Noterar att väntevärdet i denna fördelningen är precis $\mathbb{E}(X_2\mid X_1=x_1)$ och andra parametern här är precis $\mathrm{Var}(X_2\mid X_1=x_1)$.

- Allmänt:

Låt $\boldsymbol{X}\sim N(\boldsymbol{\mu},\Lambda)$ med $\det(\Lambda)>0$. Definierar en delvektor ur $\boldsymbol{X}$, säg $\boldsymbol{Y}=k$ av $\boldsymbol{X}$:s komponenter. Låt $\boldsymbol{W}=m$ av $\boldsymbol{X}$:s komponenter. det innebär att $m+k\leq n$. Det ger då $\mathbb{E}(\boldsymbol{Y}\mid\boldsymbol{W}=\boldsymbol{w})$ är normalfördelad.

Exercise 6.2 s.130: Har $\boldsymbol{X}\sim N(\boldsymbol{0},\mathbb{I}))$ och låt oss ha kovariansmatrisen

$$\Lambda=\begin{pmatrix}1&2&-1\\2&6&0\\-1&0&4\end{pmatrix}$$

Transformerar

$$Y_1=X_1+X_3$$
$$Y_2=2X_1-X_2$$
$$Y_3.=2X_3-X_2$$

- Fråga: Vilken fördelning har Y_3 givet att $Y_1=0,Y_2=1$.

Har då $\boldsymbol{Y}=B\boldsymbol{X}$ där

$$B=\begin{pmatrix}1&0&1\\2&-1&0\\0&-1&2\end{pmatrix}$$

Så hämtar bara koefficienter från transformationsvariablerna Y_i. Vet att $\boldsymbol{Y}\sim N\left(B\boldsymbol{0},B\Lambda B^T\right)$. Låt oss kalla andra parametern Λ_Y.

$$\Lambda_Y=\begin{pmatrix}3&-2&4\\-2&2&-2\\4&-2&22\end{pmatrix}$$

Ska ha

$$f_{Y_3|Y_1=0,Y_2=1}(y)=\frac{f_{\boldsymbol{Y}}(0,1,y)}{f_{Y_1,Y_2}(0,1)}\propto cf_{\boldsymbol{Y}}(0,1,y)$$

gör detta i och med att nämnaren inte har något att göra med y (här är $y=y_3$). Låt $(0,1,y)=\gamma$. Så får

$$=ce^{-\frac{1}{2}(\gamma^T\Lambda_Y^{-1}\gamma)}$$

enligt sats. Låt oss kalla det för ※ Måste invertera,

$$\Lambda^{-1}=\frac{1}{16}\begin{pmatrix}20&18&-2\\18&25&-1\\-2&-1&1\end{pmatrix}$$

så

$$\gamma^T \Lambda^{-1} = \frac{1}{16}(18 - 2y \;\; 25 - y \;\; -1 + y)$$

och vidare

$$\gamma^T \Lambda^{-1} \gamma = \frac{1}{16}(25y - y - y + y^2)$$

nu har vi 1-dim. Dvs. vi har skapat en funktion av y. Skriver

$$\gamma^T \Lambda \gamma = \left(\frac{y-1}{4}\right)^2 + \frac{24}{16}$$

Fortsättning i ※ ger

$$ce^{-\frac{1}{2}\left(\frac{y-1}{4}\right)^2}$$

Det är en 1-dim fördelning, normaltäthet. Så, slutsatsen blir att

$$Y_3 \mid Y_1 = 0, Y_2 = 1 \sim N(1, 4^2)$$

Note: Om vi inte betingar på några specifika värde så får vi istället en funktion i slutet och det blir en del av fördelning.

10. Föreläsning - Oberoende och den multivariata normalfördelningen

Innehåll: Kursboken: Kapitel 5.7-5.8, (5.9).

10.1. Kort sammanfattning från tidigare föreläsning

Multivariata normalfördelning: En slumpvektor $\boldsymbol{X}$ är multivariat normalfördelad om

$$\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{b}$$

där $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbb{I})$. Då blir

$$\mathbb{E}[\boldsymbol{X}] = A \cdot \mathbf{0} + \boldsymbol{b} = \boldsymbol{b}$$

och

$$\operatorname{Cov}[\boldsymbol{X}] = A\mathbb{I}A^T = AA^T$$

Om vi vill ha $N(\boldsymbol{\mu}, \Lambda)$, välj då

$$\boldsymbol{b} = \boldsymbol{\mu}$$
$$A = \Lambda^{\frac{1}{2}} = \mathbb{C}D^{\frac{1}{2}}\mathbb{D}^T$$

där roten fick genom att diagnoliseras kovariansmatrsen ty symmetrisk och därmed få en rotmatris.

10.1.1. Egenskaper

- $B\boldsymbol{X} + \boldsymbol{b} \sim N(\mathbf{0}, \mathbb{I})$
- Alla komponenter X_i 1-dim är normalfördelade
- Alla linjära kombinationer av X_i 1-dim
- Betingade fördelningar är normalfördelade

10.1.2. Tätheten

Definition: Om $\boldsymbol{X} \sim N(\mathbf{0}, \mathbb{I}))$ så gäller det att

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{\sqrt{\det(\Lambda)}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}^T)\Lambda^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}$$

där $\boldsymbol{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. Detta kommer ge i exponenten något polynom.

10.1.3. Momentgenererande funktioner

Definition:

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\{\boldsymbol{t}^T\boldsymbol{\mu} + t^T)\Lambda\frac{\boldsymbol{t}}{2}\}$$

där $\boldsymbol{t} = (t_1, ..., t_n) \in \mathbb{R}^n$

10.2. Oberoende och okorrelerade

Theorem: Om $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Lambda)$ och komponenterna i $\boldsymbol{X}$ oberoende så gäller det att dem är även okorrelerade. Detta är en ekvivalens, dvs. så gäller det motsatta också.

Note: $\Rightarrow$ gäller alltid. Det speciella är det motsatta, dvs. att okorrelerade $\Rightarrow$ oberoende.

Bevis på note: Räcker att visa det motsatta. Antag alltså att $\operatorname{Cov}(X_i, X_j) = 0, i \neq j$. Det betyder då att om vi tittar på kovariansmatrisen så står det 0-nollor överallt förutom huvuddiagonalen, dvs. är Λ diagonal. Det ger i sin tur att $\Lambda^{\frac{1}{2}}$ diagonal. Dvs. har vi

$$\begin{pmatrix} \sigma_1 & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sigma_n \end{pmatrix}$$

Det ger att $\boldsymbol{X} = \Lambda^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\mu}$. Z_i iid. $N(0,1)$ som ger $X_i = \sigma_i Z_i + \mu_i$. Så varje $\boldsymbol{X}$ element innehåller bara ett element ur $\boldsymbol{Z}$. Ty Z_i oberoende så gäller detta. $\square$

Så, komponenterna marginellt normalfördelade och okorrelerade räcker inte för oberoende. Krävs simultan normalfördelning.

Example: Låt $X_1 \sim N(0,1)$. Konstruerar Z oberoende av X_1 där $\mathbb{P}(Z=1) = \mathbb{P}(Z=-1) = \frac{1}{2}$ och låt även $X_2 = Z \cdot X_1$. Då: $X_2 \sim N(0,1)$. $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. Tittar på $\mathbb{E}[X_1, X_2] = \mathbb{E}[Z X_1^2] = \mathbb{E}[Z] \cdot \mathbb{E}[X_1^2] = 0$ ty symmetrisk. Det innebär alltså att dem är okorrelerade men inte oberoende.

Alltså har vi ”visat” det som nämndes ovan, vi kräver att alla komponenter är oberoende, inte bara ett komponent till exemplet här.

Exempel på varför satsen är användbar: Låt X_1, X_2 vara oberoende $N(0,1)$. Låt

$$Y_1 = X_1 + X_2, \quad Y_2 = X_1 - X_2$$

Här är $\boldsymbol{Y} = A\boldsymbol{X}$ där

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

. Så vi har att göra med en multivariat normalfördelning. Låt oss titta på $\operatorname{Cov}(Y_1, Y_2)$. Har då att

$$\begin{aligned} \operatorname{Cov}(Y_1, Y_2) &= \operatorname{Cov}(X_1 + X_2, X_1 - X_2) \\ &= \operatorname{Cov}(X_1, X_1) - \operatorname{Cov}(X_1, X_2) + \operatorname{Cov}(X_2, X_1) - \operatorname{Cov}(X_2, X_2) = 0 \end{aligned}$$

Enligt sats då så gäller det att oberoende.

Theorem: (Kommer från ett exempel ur boken) Låt X_i iid. med $N(\mu, \sigma^2)$. Då är

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

och

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

oberoende.

Proof (skiss): Inför $\boldsymbol{Y} = \left(\overline{X}, X_1 - \overline{X}_1, ..., X_n - \overline{X}_n\right)^T$. Då är

$$\boldsymbol{Y} = B\boldsymbol{X}$$

där

$$B = \begin{pmatrix} \frac{1}{n} & \dots & \dots & \dots & \frac{1}{n} \\ 1 - \frac{1}{n} & -\frac{1}{n} & \dots & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & \dots & \dots & \dots & 1 - \frac{1}{n} \end{pmatrix}$$

och eftersom $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \sigma^2\mathbb{I})$ så fås att $\boldsymbol{Y} \sim N(B\boldsymbol{\mu}, \sigma^2 BB^T)$.

$$BB^T = \begin{pmatrix} \frac{1}{n} & 0 & 0 & \dots & 0 \\ 0 & & & & \\ \vdots & & & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix}$$

Ser att Y_1 (första komponenten=$\overline{X}$) oberoende av alla andra $Y_k, k = 2, 3, ..., n+1$, följer av att en normalfördelning och att detta är en kovariansmatris. Låt $S^2 = f(Y_2, ..., Y_{n+1})$, dessa komponenter är oberoende av $\overline{X}$. $\square$

10.3. Ortogonala transformationer

Antag $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \sigma^2\mathbb{I})$, antar X_i oberoende med samma varians (godtyckligt väntevärde). Tag en ortogonal matris $\mathbb{C}$, dvs. l utgör en ortogonal bas ($\mathbb{C}\mathbb{C}^T = \mathbb{C}^T\mathbb{C} = \mathbb{I}$). Låter då $\boldsymbol{Y} = \mathbb{C}\boldsymbol{X}$. Det är en linjär transformation och därmed också N-fördelad.

$$\mathbb{E}[\boldsymbol{Y}] = \mathbb{C}\boldsymbol{\mu}$$
$$\operatorname{Cov}(\boldsymbol{Y}) = \mathbb{C}(\sigma^2\mathbb{I})\mathbb{C}^T = \sigma^2\mathbb{C}\mathbb{C}^T = \sigma^2$$

eftersom ON-matris så fås kovariansen såsom det fås. Dvs. Y_i oberoende och har samma varians σ^2. Leder till satsen:

Theorem: Om $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \sigma^2\mathbb{I})$, $\mathbb{C}$ ON-matris, då gäller det att

$$\boldsymbol{Y} = \mathbb{C}\boldsymbol{X} \sim N(\mathbb{C}\boldsymbol{\mu}, \sigma^2\mathbb{I})$$

Example: Låt X_1, X_2 iid. $N(0,1)$ och

$$Y_1 = X_1 + X_2, \quad Y_2 = X_1 - X_2$$

Då blev

$$\boldsymbol{Y} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \boldsymbol{X}$$

här har vi inte en ON-matris, nästan. Om vi multiplicerar med $\frac{1}{\sqrt{2}}$ fås en ON-bas. Så

$$\tilde{Y}_i = \frac{Y_i}{\sqrt{2}}$$

för $i = 1, 2$. Vill konstatera oberoende. Om $\tilde{Y}_i$ oberoende så gäller det att Y_i oberoende eftersom bara en multiplikation mellan dem.

$$\tilde{\boldsymbol{X}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \boldsymbol{X}$$

nu har vi en ortonormal matris. Visste att $\boldsymbol{X} \sim N(\boldsymbol{0}, \mathbb{I})$. Då fås alltså via sats ovan att $\tilde{Y}_i$ är oberoende för alla i.

Alltså, om har en ortonormal matris som transformerar X_i och X_i oberoende för alla i så gäller det även att transformationen är oberoende, dvs. alla Y_i oberoende för alla i.

Låt nu $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Lambda)$ (allmän multivariat normalfördelning). Har redan konstaterar tidigare att Λ symmetrisk som innebär det måste finnas en ON-matris $\mathbb{C}$ så att

$$\Lambda = \mathbb{C}D\mathbb{C}^T$$

där D är en diagonal matris med egenvärdena, dvs. $D = \operatorname{diag}(\lambda_1, ..., \lambda_n)$. Låt

$$\boldsymbol{Y} = \mathbb{C}^T\boldsymbol{X}$$

Då blir $\boldsymbol{Y} \sim N(\mathbb{C}^T\boldsymbol{\mu}, A)$. Alltså

$$A = \mathbb{C}^T\Lambda\mathbb{C} = \mathbb{C}^T(\mathbb{C}D\mathbb{C}^T)\mathbb{C} = D$$

ty transponanten och vanliga blir bara id-matrisen. Dvs. är komponenterna Y_i oberoende. $\operatorname{Var}(Y_i) = \lambda_i$, dvs. egenvärdena.

Theorem: Om $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Lambda)$, $\boldsymbol{Y} = \mathbb{C}^T\boldsymbol{X}$ med $\mathbb{C}$ enligt ovan. Då gäller att

$$\boldsymbol{Y} \sim N(\mathbb{C}^T\boldsymbol{\mu}, D)$$

dvs. $\{Y_i\}$ oberoende och $\operatorname{Var}(Y_i) = \lambda_i$.

11. Föreläsning - Konvergensbegrepp inom sannolikhetsteorin

Innehåll: Kursboken: Kapitel 6.1-6.3.

11.1. Repetition

11.1.1. Stora talens lag

Om $X_1, X_2, \ldots$ iid. slumpvariabler där $\mathbb{E}[X_i] = \mu, \operatorname{Var}[X_i] = \sigma^2$. Låter

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

vi har ibland även betecknat detta med S_n. Då gäller:

Stora talens lag:

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) \to 0 \quad n \to \infty, \forall \varepsilon > 0$$

Chebyshevs olikhet: Visar via Chebyshevs olikhet. Dvs:

Theorem: Om Y slumpvariabel med $\operatorname{Var}(Y) = \sigma_Y^2 < \infty$, då gäller

$$\mathbb{P}(|Y - \mu_Y| > \varepsilon) \leq \frac{\sigma_Y^2}{\varepsilon^2}$$

Låter $Y = \overline{X}_n \Rightarrow \mu_Y = \mu$ och $\sigma_Y^2 = \frac{\sigma^2}{n}$, kommer ihåg $\operatorname{Var}\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} \operatorname{Var}\left(\sum X_i\right) = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$ Detta ger då via Chebyshevs

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n} \cdot \frac{1}{\varepsilon^2} \to 0 \quad n \to \infty$$

Detta är precis vad Stora talens lag säger. □

- Vad betyder då Stora talens lag i praktik?

Om har ett tallinje med μ i mitten och $\mu \pm \varepsilon$. Det ger om vi ökar n tillräckligt mycket så kommer all massa i fördelningen tryckas in i precis μ. Det är idéen bakom konvergens i sannolikhet, all sannolikhetsmassa kommer befinna sig inuti $(\mu - \varepsilon, \mu + \varepsilon)$.

Stora talens lag är ett exempel på konvergens i sannolikhet.

11.2. Konvergens i sannolikhet

Definition: Följden X_n konvergerar i sannolikhet till slumpvariablen X då $n \to \infty$ omm $\forall \varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \quad n \to \infty$$

Notation:

$$X_n \xrightarrow{p} X$$

då $n \to \infty$

Example: Stora talens lag, $X_n = \overline{X}_n$ i detta fall och $X = \mu$ men notera att X_n inte iid.

Note: Betyder inte konvergens som i analys, dvs.

$$X_n \xrightarrow{p} X$$

betyder inte $X_n \to X$ såsom i analysmening.

11.3. (Konvergerar nästan säkert)

Definition: Konvergerar nästan säkert om

$$X_n \stackrel{a.s.}{\to} X$$

som betyder att

$$\mathbb{P}\Big(\omega : X_{n(\omega)} \to X(\omega)\Big) = 1$$

Note: Här är konvergenspilet i sannolikhetsfunktionen den pil som i analysmeningen.

Detta är ett starkare konvergens begrepp men ingår inte i kursen, kommer i Sannolikhetsteori III.

11.3.1. Centrala gränsvärdessatsen

Theorem: Har att X_i iid. med $\mathbb{E}[X_i] = \mu$ och $\mathrm{Var}[X_i] = \sigma^2$. Låter $S_n = \sum_{i=1}^n X_i$, dvs. summan upp till den n:te slumpvariabeln X. Då gäller $\forall x \in \mathbb{R}$ att

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \to \Phi(x)$$

Vi har "normerat" summan så att väntevärdet är 0 och variansen 1. Här är Φ fördelningsfunktionen för slumpvariabler som är $\sim N(0,1)$. Detta gäller för alla slumpvariabler som uppfyller dem egenskapen som nämndes ovan. Alltså:

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \approx N(0,1) \Rightarrow S_n - n\mu \approx N(0, n\sigma^2)$$

Det ger i sin tur att

$$\overline{X}_n - \mu \approx N\left(0, \frac{\sigma^2}{n}\right)$$

Så avvikelsen från snittet och väntevärdet är precis $N\left(0, \frac{\sigma^2}{n}\right)$. Så centrala gränssvärdessatsen berättar hur fort man konvergerar mot väntevärdet μ. Denna typ av konvergens definierar vi som följande.

11.4. Konvergens i fördelning

Definition: Följden X_n konvergerar i fördelning till slumpvariabeln X omm.

$$F_{X_n}(x) \to F_X(x) \qquad n \to \infty, \quad \forall x \in C(F_X)$$

där $C(F_X) = \{x : F_X(x) \text{ är kontinuerlig på } x\}$ = kontinuitetens mängd av F_X. *Dvs. var x får vara är precis alla x där F_X är kontinuerlig.*

Notation:

$$X_n \stackrel{d}{\to} X$$

då $n \to \infty$. Alternativt via

$$L(X_n) \to L(X)$$

- Se exempel 3.4 s.158 för att se varför detta krav om kontinuitet är viktig.

Note: Handlar ej om slumpvariabler utan om $F(x)$, dvs. fördelningsfunktioner och hur dem beter sig.

11.4.1. Användning

För godtyckligt stora n (beroende på kontext till exempel) gäller det att

$$\mathbb{P}(X_n \in A) \approx \mathbb{P}(X \in A) \quad A \subseteq \mathbb{R}^n$$

Example: CGS:

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \stackrel{d}{\to} Z, \qquad Z \sim N(0,1)$$

11.5. Unicitet

Theorem:

a) Antag $X_n \xrightarrow{p} X$, då är X unik. Med detta menas då: om även $X_n \xrightarrow{p} Y$ så är

$$\mathbb{P}(X = Y) = 1$$

alltså är X, Y samma.

b) Antag $X_n \xrightarrow{d} X$, då är X unik. Här betyder detta att om $X_n \to Y$ så gäller det att

$$X \stackrel{d}{=} Y$$

dvs.

$$F_X(x) = F_Y(x)$$

(notera samma utfall x). Detta gäller $\forall x C(F_X, F_Y)$

Proof:

- Börjar med a). Går in via triangelolikheten

$$|X - Y| = |X + X_n - X_n - Y| \leq |X - X_n| + |Y - X_n|$$

Så

$$\{\omega : |X - Y| > \varepsilon\} \subseteq \left\{\omega : |X_n - X| > \frac{\varepsilon}{2}\right\} \cup \left\{\omega : |X_n - Y| > \frac{\varepsilon}{2}\right\}$$

Detta kommer från triangelolikheten alltså. Detta ger då

$$\mathbb{P}(|X - Y| > \varepsilon) \leq \mathbb{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_n - Y| > \frac{\varepsilon}{2}\right) \to 0$$

Pga. våra antagande i konvergens så går högerledet mot 0.

Slutsats: $\forall \varepsilon > 0$ gäller det att $\mathbb{P}(|X - Y| > \varepsilon) = 0$ på grund av en övre begränsningen. Så

$$\mathbb{P}(|X - Y| = 0) > 0 \Rightarrow \mathbb{P}(X = Y) = 1$$

- b) x kontinuerlig punkt till F_X och F_Y. Vi "låser" x.

$$|F_X(x) - F_Y(x)| \leq \underbrace{\left|F_{X_n}(x) - F_X(x)\right|}_{\to 0 \text{ ty } X_n \xrightarrow{d} X} + \underbrace{\left|F_{X_n}(x) - F_Y(x)\right|}_{\to 0 \text{ ty } X_n \xrightarrow{d} Y}$$

Slutsats: $F_{X(x)} = F_{Y(x)}$. ☐

11.6. Sambandet mellan konvergens i sannolikhet och fördelning

Theorem:

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$$

men inte tvärtom.

Proof: (Läs s.156)

☐

Example: $X \sim Ber\left(\frac{1}{2}\right)$ med

$$X = \begin{cases} 1 & p = \frac{1}{2} \\ 0 & p = \frac{1}{2} \end{cases}$$

Inför en följd $X_1, X_2, \ldots$ där $X_n = X \quad \forall n$. Så alla X_i är lika (men är fortfarande stokastikt). Klart att

$$X_n \xrightarrow{d} X$$

Definierar $Y = 1 - X \Rightarrow Y \stackrel{d}{=} X$. Då gäller det att

$$X_n \xrightarrow{d} Y$$

Men

$$|X_n - Y| \equiv 1$$

för alltid. Så $X_n \xrightarrow{p} Y$ gäller inte.

Theorem: Om c ett reellt konstant. Och om vi har

$$X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{p} c$$

Proof: Gäller bara att visa $\Rightarrow$ hållet, det andra gäller alltid. Visar nu att $<=$ gäller. Låter $X_n \sim F_n$, då har vi att $F_n(x) \to F(x)$ som hör till konstanten c. Alltså

$$F_n(x) \to F(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases}$$

$$\mathbb{P}(|X_n - c| > \varepsilon) = 1 - \mathbb{P}(|X_n - c| \leq \varepsilon) = 1 - \mathbb{P}(c - \varepsilon \leq X_n \leq c + \varepsilon)$$

Det kan vi skriva om som

$$1 - (F_n(c + \varepsilon) - F_n(c - \varepsilon) + \mathbb{P}(X_n = c - \varepsilon)) \leq 1 - (F_n(c + \varepsilon) - F_n(c - \varepsilon))$$

Så eftersom $c + \varepsilon \geq c$ vilket ger konvergens till 1 för första termen $F_n(c + \varepsilon) \to 1$. Motsatta gäller för andra, dvs. $F_n(c - \varepsilon) \to 0$. Detta fungerar endast eftersom vi jobbar med en konstant c. $\square$

12. Föreläsning - Konvergens via transformer | STL och CGS

Innehåll: Kursboken: Kapitel 6.5-6.6.

12.1. Repetition

- $X_n \xrightarrow{p} X$ omm

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \quad n \to \infty, \forall \varepsilon$$

- $X_n \xrightarrow{d} X$ omm

$$F_{X_n}(x) \to F_X(x) \quad n \to \infty, \forall x \in C(F_X)$$

12.2. Konvergens via transformer

12.2.1. Kontinuitetsatser

Theorem: $X, X_1, X_2, \ldots$ icke-negativt heltalsvärda diskreta variabler (iochmed att vi jobbar med SGF). Om

$$g_{X_n}(t) \to g_X(t) \quad n \to \infty$$

så gäller det att

$$X_n \xrightarrow{d} X$$

Fördelen med detta är att det finns alltid en SGF, samma kan inte sägas för MGF.

Theorem: $X, X_1, X_2, \ldots$ slumpvariabler. Antag att $\exists h > 0$ så att $\psi_X(t)$ och $\psi_{X_n}(t)$ existerar för $|t| < h$ och antag att

$$\psi_{X_n}(t) \to \psi_X(t) \quad n \to \infty$$

då gäller det att

$$X_n \xrightarrow{d} X$$

Alltså, konvergens hos transformer medför konvergens i fördelning. Notera att om konvergerar i fördelning så konvergerar även i SGF, MGF men inte så jätte användbart.

Har

$$x \mapsto f(x) : x_n \to x \Rightarrow f(x_n) \to f(x)$$

samma princip gäller

$$\psi \mapsto L(\psi) : \psi_n \to \psi \Rightarrow L(\psi_n) \to L(\psi)$$

Det är härifrån namnet kontinuitetssatser kommer ifrån.

Note: $X \equiv c$, dvs. variablen är konstant, antar värde c med sannolikhet 1. Detta ger

$$\psi_X(t) = \mathbb{E}[e^{(tX]} = e^{tc}$$

$$\psi_{X_n}(t) \to e^{tc} \Rightarrow X_n \xrightarrow{d} c \Leftrightarrow X_n \xrightarrow{p} c$$

enligt tidigare sats om konstant värde där vi sa att konvergens i fördelning endast då medför konvergens i sannolikhet.

Example: Om $X_n \sim \mathrm{Bin}\left(n, \frac{\lambda}{n}\right)$. Detta är en diskret heltalsvärd variabel så jobbar med SGF.

$$\begin{aligned} g_{X_n}(t) &= (q + pt)^n \\ &= \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n} t\right)^n \\ &= \left(1 + \frac{\lambda(t-1)}{n}\right)^n \\ &= e^{\lambda(t-1)} \end{aligned}$$

Detta är då precis SGF för en Poisson variabel. Så vi har nu visat enligt ovanstående sats att

$$X_n \xrightarrow{d} X \sim \mathrm{Po}(\lambda)$$

12.3. Stora talens lag & centralgränsvärdessatsen

Theorem: $X_1, X_2, \ldots$ iid med $\mathbb{E}[X_i] = \mu < \infty$ (kräver inte ändlig varians nu längre). Skriver $\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. Då ger STL att

$$\overline{X}_n \xrightarrow{p} \mu \quad n \to \infty$$

Starka stora talens lag: Starka STL säger att

$$\overline{X}_n \xrightarrow{a.s.} \mu$$

Proof: Antag att MGF existerar (låg sannolikhet att få heltalsvärd på snittet) för $|t| < h$. Detta är ett ännu starkare krav än ändlig varians ty detta kräver att alla moment existerar (ännu bättre med $\varphi(t)$ men inte del av kursen). Det räcker alltså att visa konvergens i fördelning ty jobbar med en konstant. Alltså ska visa att $\overline{X}_n \xrightarrow{d} \mu \quad n \to \infty$.

Följer om

$$\psi_{\overline{X}_n}(t) \to e^{t\mu}$$

Låt $S_n = X_1 + \ldots + X_n$, då är

$$\psi_{\overline{X}_n}(t) = \mathbb{E}\left[e^{t\frac{1}{n}S_n}\right] = \psi_{S_n}\left(\frac{t}{n}\right)$$

enligt definitionen. Detta är en summa av oberoende slumpvariabler alltså fås då

$$\left(\psi_X\left(\frac{t}{n}\right)\right)^n$$

Noterar här att alla X_i har samma MGF igen, därmed notationen.

$$\psi_s = \mathbb{E}[E^{sX}] = \mathbb{E}\left[1 + sX + \frac{s^2X^2}{2} + \ldots\right]$$

via Taylorutveckling. Detta ger då eftersom oberoende

$$1 + s\mu + o(s)$$

där lilla ordo innehåller högre potenser, den säger något om beteendet.

Den säger att

$$\frac{o(s)}{s} \to 0 \quad s \to 0$$

Så

$$\psi_{\overline{X}_n}(t) = \left(1 + \frac{t}{n}\mu + o\left(\frac{t}{n}\right)\right)^n \to e^{t\mu}$$

($o(s)$ går mot noll väldigt snabbt). (Skriver

$$\left(1 + \frac{1}{n}\left[t\mu + \frac{o\left(\frac{t}{n}\right)}{\frac{1}{n}}\right]\right)^n$$

som kan vidare utvecklas och få den slutsatsen som vi fick.) □

Theorem: $X_1, X_2, \ldots$ iid. med $\mathbb{E}[X_i] = \mu, \mathrm{Var}[X_i] = \sigma^2 < \infty$. Skriver $S_n = X_1 + \ldots + X_n$. Då

$$\frac{S_n - n\mu}{\sqrt{\sigma^2 n}} \xrightarrow{d} N(0,1) \quad n \to \infty$$

Proof: Endast fallet då MGF existerar för $|t| < h$.

- Antag först $\mu = 0, \sigma^2 = 1$. Ska visa detta genom kontinuitetssatserna.

$$\begin{aligned} \psi_{\frac{S_n}{\sqrt{n}}}(t) &= \mathbb{E}[e^{\left(t\frac{S_n}{\sqrt{n}}\right]} \\ &= \psi_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(\psi\left(\frac{t}{\sqrt{n}}\right)\right)^n = ※ \end{aligned}$$

Via Taylor:

$$\psi(s) = 1 + s\mathbb{E}[X] + \frac{s^2}{2}\mathbb{E}[X^2] + o(s^2)$$

fortsätter från innan, noterar $\mathbb{E}[X] = 0, \mathbb{E}[X^2] = 1$,

$$※ = \left(1 + \frac{\frac{t^2}{2}}{n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{\frac{t^2}{2}}$$

samma situationen som förut.

$$\left(1 + \frac{1}{n}\left(\frac{t^2}{2} + \frac{o\left(\frac{t^2}{n}\right)}{\frac{1}{n}}\right)\right)^n \to e^{\frac{t^2}{2}}$$

Detta ger då

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

- Allmänta μ, σ^2

Hade

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)}{\sqrt{n}} = \frac{\sum_{i=1}^n \tilde{X}_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

eftersom vi har standariserad täljaren så att $\frac{X_i - \mu}{\sigma} \sim N(0,1)$. □

Exempel CH.6 problem 11: $X_n \sim \mathrm{Bin}\left(n^2, \frac{m}{n}\right)$ med $m > 0$. Visa att

$$\frac{X_n - nm}{\sqrt{nm}} \xrightarrow{d} N(0,1) \quad n \to \infty$$

Tittar på X_n som en summa av oberoende Bernoulli slumpvariabler, alltså

$$X_n \overset{d}{=} \sum_{i=1}^{n^2} V_i, \quad V_i \sim \mathrm{Be}\left(\frac{m}{n}\right)$$

har

$$\mathbb{E}[V_i] = \frac{m}{n}, \quad \mathrm{Var}[V_i] = \frac{m}{n}\left(1 - \frac{m}{n}\right)$$

Via CGS fås

$$\frac{X_n - n^2\frac{m}{n}}{\sqrt{n^2\frac{m}{n}\left(1 - \frac{m}{n}\right)}} = \frac{X_n - nm}{\sqrt{nm\left(1 - \frac{m}{n}\right)}} \to N(0,1)$$

Låt

$$W_n = \frac{X_n - nm}{\sqrt{nm}}$$

Har då

$$\frac{X_n - nm}{\sqrt{nm}} \cdot \sqrt{1 - \frac{m}{n}} \to N(0,1)$$

Alternativt löser via MGF utan CGS. Minns

$$\psi_{aY+b}(t) = e^{tb}\psi(ta)$$

Så har eftersom W_n en linjärkombination

$$W_n = \frac{1}{\sqrt{nm}} \cdot X_n - \sqrt{n}$$

detta ger då

$$\begin{aligned}\psi_{W_n}(t) &= e^{-t\sqrt{nm}}\psi_{X_n}\left(\frac{t}{\sqrt{nm}}\right) \\ &= e^{-t\sqrt{nm}}\left(\left(1 - \frac{m}{n}\right) + \frac{m}{n}e^{\frac{t}{\sqrt{nm}}}\right)^{n^2}\end{aligned}$$

om vi minns MGF för en binomial variabel. Obehagligt med e i parantesen så utvecklar via Taylor:

$$e^{\frac{t}{\sqrt{nm}}} = 1 + \frac{t}{\sqrt{nm}} + \frac{t^2}{2nm} + o\left(\frac{1}{n}\right)$$

så fortsättningsvis

$$\begin{aligned}&e^{-t\sqrt{nm}}\left(1 - \frac{m}{n} \cdot \frac{m}{\sqrt{nm}} + \frac{\frac{t^2}{2}}{n^2} + o\left(\frac{1}{n^2}\right)\right)^{n^2} \\ &= e^{-t\sqrt{nm}}\left(1 - \frac{1}{n^2}\left(\sqrt{nm} \cdot t + \frac{t^2}{2} + \frac{o\left(\frac{1}{n^2}\right)}{\frac{1}{n^2}}\right)\right)^{n^2} \\ &\sim e^{\sqrt{nm}t + \frac{t^2}{2} + \frac{o\left(\frac{1}{n^2}\right)}{\frac{1}{n^2}}}\end{aligned}$$

detta går då vid insättning av gränsen

$$e^{-t\sqrt{nm}}\left(1 - \frac{1}{n^2}\left(\sqrt{nm} \cdot t + \frac{t^2}{2} + \frac{o\left(\frac{1}{n^2}\right)}{\frac{1}{n^2}}\right)\right)^{n^2} \to e^{\frac{t^2}{2}}$$

Alltså har vi visat via MGF att $W_n \xrightarrow{d} N(0,1)$.

2

13. Föreläsning - Mer om konvergens.

Innehåll: Kursboken: Kapitel 6.7.

13.1. Konvergens för summor/differanser/kvoter...

- $X_n \to X$ och $Y_n \to Y$, betyder det då att $X_n + Y_n \to X + Y$. Detta är dagens frågenställning. Svaret kommer bero på typen av konvergens.

Theorem: Om $X_n \xrightarrow{p} X$ och $Y_n \xrightarrow{p} Y$ så gäller det att

$$X_n + Y_n \xrightarrow{p} X + Y$$

Note: Kräver inte något om oberoende.

Theorem: Om $X_n \xrightarrow{d} X$ och $Y_n \xrightarrow{d} Y$ och antar att X, Y oberoende och även $\forall n : X_n, Y_n$ oberoende. Då gäller det att

$$X_n + Y_n \xrightarrow{d} X + Y$$

Note: Vi kräver alltså oberoende här eftersom konvergens i fördelning är ett svagare krav.

Theorem: Om Om $X_n \xrightarrow{d} X$ och $Y_n \xrightarrow{p} a$ för något konstant a så gäller det att

$$X_n + Y_n \xrightarrow{d} X + a$$

Note: Dessa satser gäller för vilka operatorer som helst. T.ex. tredje satsen blir då

$$\begin{aligned} &X_n - Y_n \xrightarrow{d} X - a \\ &X_n Y_n \xrightarrow{d} aX \\ &\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{a} \quad a \neq 0 \end{aligned}$$

Proof:

Sats 1. Enligt definition vill vi visa

$$\mathbb{P}(|X_n + Y_n - (X + Y)| > \varepsilon) \to 0 \quad n \to \infty$$

Via triangeolikheten har vi då

$$\mathbb{P}(|X_n + Y_n - (X + Y)| > \varepsilon) \leq \mathbb{P}\left(\left\{|X_n - X| > \frac{\varepsilon}{2}\right\} \cup \left\{|Y_n - Y| > \frac{\varepsilon}{2}\right\}\right)$$

Eftersom union så tänk på "eller" att något av dem två händelser sker. Har vidare

$$\leq \mathbb{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|Y_n - Y| > \frac{\varepsilon}{2}\right)$$

vill visa att detta går mot 0 då $n \to \infty$. Eftersom $X_n \to X, Y_n \to Y$ så fås konvergens till 0 ty differansen blir 0.

Sats 2. Antag MGF existerar (om du vill ha starkare krav så jobba med karakteristiska) då $\exists h > 0$ och $|t| < h$. På grund av konvergens i fördelning så har vi även konvergens i MGF. Dvs. om $X_n \xrightarrow{d} X \Leftrightarrow \psi_{X_n}(t) \to \psi_X(t)$. Detsamma för Y följden. Det förenklar satsens bevis en del.

$$\psi_{X_n + Y_n}(t) = \psi_{X_n}(t)\psi_{Y_n}(t)$$

ty oberoende. Vet att

$$\begin{aligned} \psi_{X_n}(t)\psi_{Y_n}(t) &\to \psi_X(t)\psi_Y(t) \\ &= \psi_{X+Y}(t) \end{aligned}$$

återigen på grund av oberoende. Detta visar sats 2.

Sats 3 (skiss). Se Gut p.167 för hela beviset. Idén är att använda oss av

$$X_n + Y_n \leq x$$

och

$$|Y_n - a| \leq \varepsilon$$

om n stort. Grundidén är att

$$X_n \leq x - Y_n \leq x - (a - \varepsilon) = x - a + \varepsilon$$

Detta ger då

$$\limsup F_{X_n+Y_n}(x) \leq F_X(x-a) = F_{X+a}(x) \quad \varepsilon \geq 0$$

ty $X \leq x - a \Leftrightarrow X + a \leq x$. Motsvarande undre gräns kan fås via andra håller

$$\liminf F_{X_n+Y_n}(x) \geq F_{X+a}(x), \quad \varepsilon \geq 0$$

□

Example: Låt $\{X_i\}$ vara iid. följd av $U(0,1)$ slumpvariabler. Visa att

$$\frac{X_1 + \ldots + X_n}{X_1^2 + \ldots + X_n^2} \xrightarrow{p} \frac{3}{2}$$

När man får något som att visa konvergens mot konstant och innehåller summor så är det bra att titta på Stora talens lag. Låt oss titta på summorna separat:

$$\frac{1}{n}(X_1 + \ldots + X_n) \xrightarrow{p} \mathbb{E}[X_1] = \frac{1}{2}$$

och

$$\frac{1}{n}(X_1^2 + \ldots + X_n^2) \xrightarrow{p} \mathbb{E}[X_1^2] = \int_0^1 x^2 1 dx = \frac{1}{3}$$

enligt STL. Nu vet vi täljarens och nämnarens beteende. Klistrar ihop med hjälp av sats 1 (eftersom många konvergens i sannolikheter) och får

$$\frac{\frac{X_1+\ldots+X_n}{n}}{\frac{X_1^2+\ldots+X_n^2}{n}} \xrightarrow{p} \frac{\frac{1}{2}}{\frac{1}{3}} = \frac{3}{2}$$

och därmed har vi visat konvergens.

Example: Har $\{Y_i\}$ iid. följd med $U(-1,1)$. dvs. $\mathbb{E}[Y] = 0, \operatorname{Var}[Y] = \frac{1}{3}$. Visa att

$$\frac{\sum_{i=1}^n Y_i}{\sqrt{n} \max_{1 \leq i \leq n} Y_i} \xrightarrow{d} N\left(0, \frac{1}{3}\right)$$

Om du ser en normalfördelning så tänk CGS. Har

$$\frac{\sum Y_i}{\sqrt{n}} = \sqrt{\frac{1}{3}} \frac{\sum Y_i}{\sqrt{\frac{1}{3}n}} \xrightarrow{d} \sqrt{\frac{1}{3}} N(0,1) = N\left(0, \frac{1}{3}\right)$$

enligt CGS. För att få detta måste vi lösa max i nämnaren men eftersom likformigt på (-1,1) så får vi direkt att $\max Y_i = 1$. Dvs. ska visa att

$$\begin{aligned}
\mathbb{P}(|\max Y_i - 1| > \varepsilon) &= \mathbb{P}(\max Y_i < 1 - \varepsilon) \\
&= \mathbb{P}(\text{alla } Y_i < 1 - \varepsilon) \\
&= \mathbb{P}(Y_1 < 1 - \varepsilon)^n \\
&= \left(\int_{-1}^{1-\varepsilon} \frac{1}{2} dx\right)^n \\
&= \left(\frac{2-\varepsilon}{2}\right)^n \to 0
\end{aligned}$$

Så då har vi visat att $\max Y_i \xrightarrow{p} 1$. Då ger sats 3, ty konvergens i fördelning och konvergens i sannolikhet mot en konstant:

$$\frac{\sum_{i=1}^n \frac{Y_i}{\sqrt{n}}}{\max Y_i} \xrightarrow{d} N\left(0, \frac{1}{3}\right)$$

där täljaren konvergar i fördelning $N\left(0, \frac{1}{3}\right)$ och nämnaren konvergerar i sannolikhet mot 1.

13.2. Konvergens för funktioner av slumpvariabler

- Tittar på fallet $X_n \xrightarrow{p} a$ (kan konvergera mot slumpvariabel, eller konvergerar i fördelning eller annat också), ger detta då att $g(X_n) \xrightarrow{p} g(a)$?

Theorem: Har $X_n \xrightarrow{p} a$, givet att $g(x)$ är kontinuerlig i punkten a så gäller det att

$$g(X_n) \xrightarrow{p} g(a)$$

Proof: Enligt definition, tag $\varepsilon > 0$, ska visa att

$$\mathbb{P}(|g(X_n) - g(a)| > \varepsilon) \to 0$$

Att vara kontinuerlig i punkten a betyder att $\exists \delta > 0$ så att

$$|x - a| \leq \delta \Rightarrow |g(x) - g(a)| \leq \varepsilon$$

mer eller mindre definitionen av kontinuitet. Alltså gäller det att

$$|g(x) - g(a)| > \varepsilon \Rightarrow |x - a| > \delta$$

Funktionsvärde långt isär betyder att argumenten ligger långt isär i princip. Så

$$\mathbb{P}(|g(X_n) - g(a)| > \varepsilon) \leq \mathbb{P}(|X_n - a| > \delta)$$

enligt ovan. När $n \to \infty$ då fås konvergens mot 0 ty $X_n \xrightarrow{p} a$. □

Tenta 2004, exempel: Har $\{X_i\}$ iid. med

$$\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = \frac{1}{2}$$

och har även $\{Y_i\}$ iid. med

$$\mathbb{P}(Y_i = 0) = \mathbb{P}(Y_i = 1) = \frac{1}{2}$$

Visa att

$$\frac{\sum X_i \cdot \sqrt{\sum Y_i}}{\sum X_i + \sum Y_i} \xrightarrow{d} \quad ?$$

Bestäm gränsfördelningen alltså. Om vi börjar med summan av X_i, vet att för alla $\mathbb{E}[X] = 0$ och $\mathrm{Var}[X] = 1$

$$\frac{\sum X_i}{\sqrt{n}} \xrightarrow{d} N(0,1)$$

enligt CGS. Går vidare med täljaren.

$$\frac{\sqrt{\sum Y_i}}{\sqrt{n}} = \sqrt{\frac{\sum Y_i}{n}} \to \sqrt{\mathbb{E}[Y_i]} = \sqrt{\frac{1}{2}}$$

enligt STL och $g(x) = \sqrt{x}$ är kontinuerlig i $x = \frac{1}{2}$. Låt hela bråket vara W_n. Så har

$$W_n = \frac{\frac{\sum X_i}{\sqrt{n}} \cdot \left(\frac{\sqrt{\sum Y_i}}{\sqrt{n}}\right)}{\frac{\sum X_i}{n} + \frac{\sum Y_i}{n}} \to \frac{N(0,1) \cdot \frac{1}{\sqrt{2}}}{\mathbb{E}[X_i] + \mathbb{E}[Y_i]} = \frac{\frac{1}{\sqrt{2}} N(0,1)}{0 + \frac{1}{2}} = \sqrt{2} N(0,1) = N(0,2)$$