Theory of Statistics - Complete Notes

Hampus Beijer

August 16, 2025

Introduktion

Dessa anteckningar täcker hela kursen Theory of Statistics I, inklusive alla föreläsningar från 1 till 14. Materialet innehåller grundläggande begrepp i sannolikhetsteori, statistisk inferens, Bayesiansk analys och hypotesprövning.

1 Föreläsning 1: Grundläggande sannolikhetsbegrepp

1.1 Mängdlära

- **Utfallsrum (S)**: Mängden av alla möjliga utfall
- **Händelse**: En delmängd av utfallsrummet
- **Operationer**:
  - Union $A \cup B$: Utfall i A eller B eller båda
  - Snitt $A \cap B$: Utfall i både A och B
  - Komplement A^c: Utfall inte i A
- **De Morgans lagar**:
$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

1.2 Sannolikhetstolkningar

- **Frekventistisk**: Långsiktig relativ frekvens
$$P(A) = \lim_{n\to\infty} \frac{\text{Antal gånger A inträffar}}{n}$$
- **Bayesiansk**: Grad av tro
- **Axiom**:

1. $P(A) \geq 0$ för alla händelser A
2. $P(S) = 1$
3. För disjunkta händelser $A_1, A_2, \ldots$:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

1.3 Betingad sannolikhet

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{om } P(B) > 0$$

1.4 Bayes sats

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

För en partition $B_1, \ldots, B_k$:

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{k} P(A \mid B_j)P(B_j)}$$

1.5 Kombinatorik

- **Permutationer**: Ordade arrangemang

$$P(n, k) = \frac{n!}{(n-k)!}$$

- **Kombinationer**: Oordade urval

$$C(n, k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- **Multinomial**:

$$\binom{n}{n_1, \ldots, n_k} = \frac{n!}{n_1! \cdots n_k!}$$

2 Föreläsning 2: Slumpvariabler,

Föreläsningen täcker:

- Slumpvariabler (SV)
- Diskreta fördelningar (PF)
- Kontinuerliga fördelningar (PDF)
- Kumulativ fördelningsfunktion (CDF)

2.1 Stokastiska variabler (SV)

- **Definition**: En funktion $X : S \to \mathbb{R}$ som avbildar utfall på reella tal
- **Exempel**: Tärningskast

$$X(s) = \begin{cases} 0 & \text{om } s \in \{1,2,3\} \\ 1 & \text{om } s \in \{4,5,6\} \end{cases}$$

2.2 Fördelingstyper

En SV är diskret om den endast kan ta ett ändligt eller räknebart antal olika värden $x_1, x_2, ...$ En slumpvariabel är kontinuerlig om den kan ta varje värde inom ett givet intervall. För kontinuerliga och diskreta sannolikhetsfördelningar finns följande defitioner:

- **Diskret (Ändligt/uppräkneligt stöd.)**: En diskret variabel har en sannolikhetsförunktion som på engelska är **probablitiy function, p.f**

$$f(x) = P(X = x)$$

- **Kontinuerlig**: Sannolikhetsfunktionen för en kontinuerlig variabel kallas för en **Probablity Density Function, PDF** eller på svenska **Täthetsfunktion**, och är som ett histogram med små bin widths. Täthetsfunktion defineras:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Täthetsfunktionen för en kontinuerlig variabeö har följand egenskaper:

- $P(a \leq X \leq b) = \int_a^b f(x)dx$
- $f(x) \geq 0$ för alla x
- $\int_a^b f(x)dx = 1$

2.3 Viktiga fördelningar

- **Likformig diskret** ($X \sim U(a,b)$):

$$f(x) = \begin{cases} \frac{1}{b-a+1} & a, .., b \\ 0 & \text{annars} \end{cases}$$

- **Bernoulli** ($X \sim \text{Bern}(p)$):

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

- **Binomial** ($X \sim \text{Bin}(n,p)$):

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

- **Likformig kontinuerlig** ($X \sim U(a,b)$):

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{annars} \end{cases}$$

- **Normal** ($X \sim N(\mu, \sigma^2)$):

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

2.4 Kumulativ fördelningsfunktion (CDF)

Kumulativa fördelningsfunktionen eller **Cumulative Distribution Function (CDF)** för en kontinuerligt och diskret slumpvariabel $F(x)$ defineras:

$$F(x) = P(X \le x), -\infty \le x \le \infty$$

En CDF har följande egenskaper

- Icke-avtagande: Om $x_1 \le x_2$ då är $F(x_1) \le F(x_2)$
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$
- Alltid kontinuerlig från höger $F(x) = F(x^+) = P(X \le x)$
- Interval sannolikheter: $P(x_1 < X \le x_2) = F(x_2) - F(x_1)$
- Skillnaden mellan strikta och svaga olikheter är viktiga $P(X < x) = F(x^-)$ och $P(X \le x) = F(x) = F(x^-) + P(X = x)$

Kontinuerliga variabler har en **relation mellan pdf och cdf**:

$$F(x) = \int_{-\infty}^{x} f(t)dt \text{ och omvänt gäller} \frac{dF(x)}{d} = f(x)$$

3 Föreläsning 3: Funktioner av stokastiska variabler

3.1 Kvantiler

1: Definition av p-kvantil: p-kvantilen $F^{-1}(p)$ är det minsta x så att $F(x) \ge p$. Denna definiton fungerar för alla olika typer av CDF:er även om de är diskontinuerliga eller 1:1 över hela stödet för X.

2: Specialfall för kontinuerlig CDF Om CDFn, $F(x)$ är kontinuerlig och 1:1 (dvs invers finns) så är p-kvantilen helt enkelt inversen av CDFn $x = F^{-1}(p =)$. Här kan man direkt lösa p = F(x):

$$x = F^{-1}(p)ochp = F(x) \tag{1}$$

p är sannolikheten mellan 0 och 1 vilket representerar den kumulativa fördelningen upp till värdet x. Det är andelen av fördelningen som ligger till vänster om x. Exempelvis, p=0.975 innebär att 97.5% av data $\leq x$. +x är kvantilvärden eller gränsvärdet i fördelningen som motsvarar sannolikheten p. Exempel med median:

Vi har en SV $X \sim exp(\lambda)$ som har täthetsfunktionen $f(x) = \lambda e^{-\lambda x}, x > 0$. Konsekvent är då CDF $F(x) = 1 - e^{-\lambda x}$. **Vi söker medianen m = 0.5**. Teorin säger då att vi kan finna denna genom $F(m = 0.5) = 1 - e^{-\lambda x} = 0.5$. Lös sedan ut **m** och då fås $+e^{-\lambda m} = -0.5$ så $e^{-\lambda m} = 0.5$ sen $-\lambda m = ln(1/2)$ som kan skrivas som $-\lambda m = -ln(2)$ och $m = ln(2)/\lambda$. **3:Kända kvantiler**

- Median: $F^{-1}(0.5)$
- Första kvantil: $F^{-1}(0.25)$
- Tredje kvantil: $F^{-1}(0.75)$

3.2 Funktioner av en diskret SV

För att förstå funktioner av slumpvariabler. Föreställ dig följande scenario.

1:Situationen Vi har fördelningen för X t.ex sannolikhetsfunktionen för f(x) är känd. Fördelningen för Y=r(X) där r(.) ären given funktion (linjär, invertering, log, logit osv). Målet är då att finna funktionen av SV Y som är en funktion av SV X, Y = r(X).

2: Transformeringsexempler

- $Y = a + bX$ (Linjär transformation)
- $Y = 1/X$ (Invertering)
- $Y = ln(X)(Logaritm)$
- $Y = log(X/1 - X)(Logit - transform)$

3. Metod för att hitta sannolikhetsfunktion (PF) för Y=r(X) när X är en diskret SV:

$$g(y) = P(Y = y) = P[r(X) = y] = \sum_{x:r(x)=y} = f(x) \tag{2}$$

där,

- $g(y)$ är sannolikhetsfunktionen för Y. Dvs, den anger sannolikheten att Y antar värdet y

- $P[r(X) = y]$ är sannolikheten att transformationen $r(X)$ resulterar i värdet y
- $r : r(x) = y$ innebär alla x-värden som uppfyller $r(x) = y$
- f(x) är sannolikheten att $X = x$
- Summationen $\sum f(x)$ är summan av alla f(x) för dessa x värden är $g(y)$

Diskret fall, exempel: Antag att X är ett tärningskast med $X \in 1, 2, 3, 4, 5, 6$ med likformig fördelning, $f(x) = 1/6$. Låt $Y = r(X) = X$. Då finns det två olika möjligtheter (Y=0 eller Y=1). $g(y)$ kan du beräknas som Y=1 eller Y=1 enligt:

- $g(0) = \sum_{x:xjämt} = f(2) + f(4) + f(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$
- $g(1) = \sum_{x:xjämt} = f(1) + f(3) + f(5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

3.3 Funktion av kontinuerligt SV

När man arbetar med en kontinuerlig SV är stegen annrolunda för att finna fördelningen, så hur hittar man då fördelningen för Y=r(X) då X är känd?

1: Förutsättningar Funktionen $Y = r(x)$ måste vara en 1:1 funktion (inverterbar)

2: CDF för Y CDFn för Y betecknas $G(y)$ och skrivs som:

$$G(y) = P(Y \leq y) = P[r(X) \leq y] = P[X \leq s(y)] = \int_a^{s(y)} f(x)dx = F[s(y)] - F[a] \tag{3}$$

där F(X) är CDFn för X.

3: PDF för Y Derivera G(y) med avseende på y då får man täthetsfunktionen (PDF) för Y:

$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right|, \alpha < y < \beta \tag{4}$$

där $s(y) = r^{-1}(y)$. Denna metod är förutsätter att inversen exiterar. Metoden kallas för variablentransformation och använder kedjeregeln för att besvara sannolikhet:

4 Föreläsning 4: Bivariata och multivariata fördelningar

4.1 Bivariat fördelning

En bivariat fördelning är en samling av alla sannolikheter för händelser där både X och Y är iblandade, dvs $Pr[(X, Y) \in C]$ där C är en delmängd av möjliga utfall (t.ex X=0 och Y=¿ 1). Riktiga exempler på detta kan vara: Antal läkarbesök X och akutbesök Y, där C kan vara scenarier som ”inga läkarbesök men minst ett akutbesök”.

4.2 Joint Probablity Function: Diskreta variabler och gemensam sannolikhetsfunktion

För diskreta variabler defineras den gemensamma sannolikhetsfunktionen:

$$f(x, y) = P(X = x, Y = y) \tag{5}$$

Sannolikheten för händelse C beräknas då genom att summera $f(x, y)$ över alla möjliga utfall i C:

$$Pr[(X, Y) \in C] = \int \int_C f(x, y) = \sum_{(x,y) \in C} f(x, y) \tag{6}$$

Då finns följande normeringsvillkor som säger att summan av alla möjliga sannolikheter måste summeras till 1:

$$\sum_{All x,y} f(x, y) = \sum_x \sum_y f(x, y) = 1 \tag{7}$$

Sammanfattningsvis kan man säga att den bivariata fördelningen beskriver hur X och Y beter sig tillsammans.

4.3 Continuous joint distributions: Kontinuerlig bivariat fördelning

Den kontinuerliga bivariata fördelningen (joing pdf) deineras enligt:

$$Pr[(X, Y) \in C] = \int \int_C f(x, y) \tag{8}$$

Där $f(x, y) \geq 0$ är den gemensamma täthetsfunktionen för X,Y.

- För en skild variabler (unvariat) motsvaras sannolikheten av area under täthetskurvan.
- För två varaibler **bivariat** motsvarar sannolikheten av volymen under täthetsytan över regionen C, även kallas bottenplattan.
- Den totala volymen under täthetsytan 'r alltid 1: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$

4.4 Joint CDF

4.4.1 1: Gemensam fördelningsfunktion, joint cdf

4.4.2 2: Sannolikheter för rektangulära områden

4.5 Gemensam fördelning

- **Gemensam CDF**:

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

- **Gemensam pdf** (kontinuerlig):

$$P((X,Y) \in C) = \iint_C f_{X,Y}(x,y)dxdy$$

4.6 Marginalfördelningar

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

4.7 Betingade fördelningar

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

4.8 Oberoende

X och Y är oberoende om:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

4.9 Bivariat normal

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$

Betingad fördelning:

$$X|Y = y \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1-\rho^2)\right)$$

5 Föreläsning 5: Väntevärde och varians

5.1 Väntevärde

$$E[X] = \begin{cases} \sum_x xf(x) & \text{diskret} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{kontinuerlig} \end{cases}$$

5.2 Varians

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - \mu^2$$

5.3 Kovarians och korrelation

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

5.4 Momentgenererande funktion

$$M_X(t) = E[e^{tX}]$$

Moment:

$$E[X^n] = M_X^{(n)}(0)$$

6 Föreläsning 6: Kontinuerliga fördelningar

6.1 Gammafördelning

$$X \sim \text{Gamma}(\alpha, \beta)$$

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

$$E[X] = \alpha/\beta, \quad \text{Var}(X) = \alpha/\beta^2$$

6.2 Betafördelning

$$X \sim \text{Beta}(\alpha, \beta)$$

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

$$E[X] = \frac{\alpha}{\alpha+\beta}$$

7 Föreläsning 7: Gränsvärdessatser

7.1 Markovs olikhet

För $X \geq 0$:

$$P(X \geq t) \leq \frac{E[X]}{t}$$

7.2 Chebyshevs olikhet

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

7.3 De stora talens lag

För oberoende likafördelade X_i med $E[X_i] = \mu$:

$$\bar{X}_n \xrightarrow{p} \mu$$

7.4 Centrala gränsvärdessatsen

För oberoende likafördelade X_i med $E[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

8 Föreläsning 8: Simuleringsmetoder

8.1 Monte Carlo-integration

$$E[g(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

8.2 Invers transformationsmetod

För $U \sim U(0,1)$, sätt $X = F^{-1}(U)$.

9 Föreläsning 9-10: Bayesiansk inferens

9.1 Bayes sats

$$p(\theta \mid x) = \frac{f(x \mid \theta)p(\theta)}{f(x)} \propto \mathcal{L}(\theta; x)p(\theta)$$

9.2 Konjugerade priorer

- Binomial - Beta prior
- Normal - Normal prior (känd varians)
- Poisson - Gamma prior

9.3 Prediktiv posterior

$$p(x_{ny} \mid x) = \int f(x_{ny} \mid \theta)p(\theta \mid x)d\theta$$

10 Lecture 10: Bayesian Analysis

10.1 Non-informative Priors

- Bernoulli model: $\theta \sim \mathrm{Beta}(1,1)$ (uniform prior)
- Normal model: $p(\theta) = 1$ (improper prior)

$$\int p(\theta)d\theta = \infty \quad \text{(must verify posterior is proper)}$$

10.2 Normal Data Analysis

- Known variance, uniform prior:

$$\theta|x \sim N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

- Known variance, normal prior $\theta \sim N(\mu_0, \tau_0^2)$:

$$\begin{aligned}\frac{1}{\tau_n^2} &= \frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\\ \mu_n &= w\bar{x} + (1-w)\mu_0\\ \text{where } w &= \frac{n/\sigma^2}{n/\sigma^2 + 1/\tau_0^2}\end{aligned}$$

10.3 Poisson Model

$$\begin{aligned}\text{Likelihood: } & p(y|\theta) \propto \theta^{\sum y_i} \exp(-\theta n)\\ \text{Conjugate prior: } & \theta \sim \text{Gamma}(\alpha, \beta)\\ \text{Posterior: } & \theta|y \sim \text{Gamma}\left(\alpha + \sum y_i, \beta + n\right)\end{aligned}$$

11 Lecture 11: Estimation Methods

11.1 Bayes Estimators

$$\begin{aligned}\text{Quadratic loss: } & L(a,\theta) = (\theta - a)^2 \Rightarrow \text{Posterior mean}\\ \text{Linear loss: } & L(a,\theta) = |\theta - a| \Rightarrow \text{Posterior median}\end{aligned}$$

11.2 MLE Examples

- Poisson: $\hat{\theta} = \bar{x}$
- Bernoulli: $\hat{\theta} = \bar{x}$
- Normal (known σ^2): $\hat{\mu} = \bar{x}$

11.3 Method of Moments

Solve system:

$$\mu_j(\theta) = m_j \quad \text{for } j = 1, ..., k$$

12 Lecture 12: Sampling Distributions

12.1 Chi-Square Distribution

- $X \sim \chi^2_m$ is $Gamma(m/2, 1/2)$
- Properties:

$$E(X) = m$$
$$Var(X) = 2m$$

- Sum of squares:

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{(unknown } \mu\text{)}$$

12.2 t-Distribution

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

13 Lecture 13: Confidence Intervals

13.1 Normal Mean CI

$$\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$$

13.2 Normal Variance CI

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right)$$

13.3 Fisher Information

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X|\theta)\right]$$
$$I_n(\theta) = nI(\theta)$$

14 Lecture 14: Hypothesis Testing

14.1 Test Statistics

- Normal (known σ^2):

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

- t-test (unknown σ^2):

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

- F-test (variance comparison):

$$V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)} \sim F_{m-1,n-1}$$

14.2 Likelihood Ratio Test

$$\Lambda = \frac{f(x|\hat{\theta}_0)}{f(x|\hat{\theta})}, \quad -2\ln\Lambda \overset{a}{\sim} \chi_k^2$$

14.3 Error Types

Type I	Reject H_0 when true
Type II	Fail to reject H_0 when false