



Föreläsning 04

Statistisk inläring och dataanalys (Kungliga Tekniska Högskolan)

Föreläsning 4

Nu har vi byggt upp en stor samling av statistiska modeller (båda endimensionella och flerdimensionella) som kan bero på en eller flera parametrar $\theta_1, \dots, \theta_d$. I praktiken får vi inte dessa parametrar men istället bara data vilka vi skulle vilja använda för att skatta parametrarna. Ett första steg i denna skattning är att förstå hur datan genereras från en viss fördelning förhåller sig till själva fördelningen. Ett vanligt antagande är att datan är ett stickprov x_1, \dots, x_n ; det vill säga, en sekvens av utfall där $X_1 = x_1, \dots, X_n = x_n$ och X_1, \dots, X_n är oberoende och likafördelade.

Det följer från oberoende av X_1, \dots, X_n att den simultana fördelningen av $\mathbf{X} = (X_1, \dots, X_n)$ är relativt enkel:

$$f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \prod_{i=1}^n f_{X_i}(x_i).$$

Denna fördelning beror ofta på några parametrar $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$; dvs.

$$f_{\mathbf{X}}(x_1, \dots, x_n | \boldsymbol{\theta}) = \prod_{i=1}^n f_{X_i}(x_i | \boldsymbol{\theta}).$$

Exempel 4.1. Låt X_1, \dots, X_n vara brinntiderna för n oberoende glödlampor. Vi antar att X_1, \dots, X_n är likafördelade med sannolikhetsfunktion

$$f_X(x) = \lambda e^{-\lambda x}, \quad 0 \leq x < \infty, \quad \lambda > 0;$$

dvs. X_1, \dots, X_n är $\text{Exp}(\lambda)$ -fördelade. Eftersom X_1, \dots, X_n är oberoende har vi

$$f_{\mathbf{X}}(x_1, \dots, x_n | \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} = \lambda^n e^{-n\lambda \bar{x}},$$

där $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ är stickprovsmedelvärdet av x_1, \dots, x_n . Vi kan använda denna simultana fördelning för att beräkna vissa sannolikheter; exempelvis, sannolikheten att alla glödlamporna håller längre än 2 år:

$$\begin{aligned} P(X_1 > 2, \dots, X_n > 2) &= P(X_1 > 2) \cdots P(X_n > 2) && \text{(oberoende),} \\ &= P(X > 2)^n && \text{(likafördelade),} \\ &= \left(\int_2^\infty \lambda e^{-\lambda x} dx \right)^n, \\ &= (e^{-2\lambda})^n, \\ &= e^{-2n\lambda}. \end{aligned}$$

4.1 Statistikor

I Exempel 4.1 skrev vi simultan fördelningen av X_1, \dots, X_n som en funktion av \bar{x} – ett värde som sammanfattar datan x_1, \dots, x_n :

$$f_{\mathbf{X}}(x_1, \dots, x_n | \lambda) = \lambda^n e^{-n\lambda \bar{x}}.$$

Det är ofta användbart att sammanfatta ett stickprov. Detta görs med hjälp av en funktion

$$T: \mathbb{R}^n \longrightarrow \mathbb{R}^m,$$

där $T(X_1, \dots, X_n) = T(\mathbf{X}) = \mathbf{Y}$ som kallas en *statistika*.

Exempel 4.2. Några vanliga statistikor är

- *stickprovsmedelvärde*: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$,
- *stickprovsvariansen*: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, och
- *stickprovsstandardavvikelsen*: $S = \sqrt{S^2}$.

Om vi har ett stickprov x_1, \dots, x_n skriver vi $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ respektiv $s = \sqrt{s^2}$.

Eftersom T är en funktion av stokastiska variabler är $\mathbf{Y} = T(\mathbf{X})$ bara en transformation av stokastiska variabler. Det följer att en statistika $T(\mathbf{X})$ har en fördelning. Denna fördelning kallas *samplingfördelningen*. Samplingfördelningen för statistikor som stickprovsmedelvärde och stickprovsvariansen har bra egenskaper:

Sats 4.1. X_1, \dots, X_n oberoende och likafördelade med väntevärde och varians μ respektiv $\sigma^2 < \infty$ har vi att

1. $E[\bar{X}] = \mu$,
2. $\text{Var}[\bar{X}] = \frac{\sigma^2}{n}$, och
3. $E[S^2] = \sigma^2$.

Bevis. För (a) har vi att

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} n E[X_1] = \mu,$$

där vi använder att $E[X_1] = E[X_i]$ för alla $i = 1, \dots, n$ eftersom X_1, \dots, X_n är likafördelade.

För (b) har vi att

$$\text{Var}[\bar{X}] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} n \text{Var}[X_1] = \frac{\sigma^2}{n},$$

där vi använder först att X_1, \dots, X_n är oberoende och då att de är likafördelade.

För (c) observerar vi först att

$$\begin{aligned} \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n (X_i - \bar{X} + \bar{X})^2, \\ &= \sum_{i=1}^n ((X_i - \bar{X})^2 + 2\bar{X}(X_i - \bar{X}) + \bar{X}^2), \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2\bar{X} \sum_{i=1}^n X_i - \sum_{i=1}^n 2\bar{X}^2 + \sum_{i=1}^n \bar{X}^2, \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + 2n\bar{X}^2 - 2n\bar{X}^2 + n\bar{X}^2, \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n\bar{X}^2. \end{aligned}$$

Därför har vi att

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right).$$

Det följer att

$$\begin{aligned} E[S^2] &= E\left[\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)\right], \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E[X_i^2] - n E[\bar{X}^2] \right), \\ &= \frac{1}{n-1} (n E[X_1^2] - n E[\bar{X}^2]), \\ &= \frac{1}{n-1} \left(n(\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \right), \end{aligned}$$

som visar att $E[S^2] = \sigma^2$. □

Sats 4.1 säger att \bar{X} respektive S^2 är bra skattningar för μ och σ^2 . Vi ska se senare att ekvivalenserna $E[\bar{X}] = \mu$ och $E[S^2] = \sigma^2$ betyder att skattningarna \bar{X} respektiv S^2 är *väntevärdesriktiga skattningar* för μ och σ^2 . Medan det kan verka konstigt att dela med $n - 1$ i den tredje formeln i Sats 4.1 ska vi se att skattningen inte skulle vara väntevärdesriktig om vi hade dividerat med n istället.

Om X_1, \dots, X_n är oberoende och likafördelade med täthetsfunktion (eller sannolikhetsfunktion) $f_X(x)$ i en läge-skalafamilj med standardtäthetsfunktion $f_Z(z)$ för vilken det är lättare att beräkna samplingfördelningen för \bar{Z} kan vi få direkt en formel för samplingfördelningen för \bar{X} . Eftersom $f_X(x)$ är i läge-skalafamiljen med standardtäthetsfunktion $f_Z(z)$ har vi att $X_i = \sigma Z_i + \mu$ för några μ och $\sigma > 0$ och Z_i som är $f_Z(z)$ -fördelade. Eftersom X_1, \dots, X_n är oberoende följer det även att Z_1, \dots, Z_n är oberoende. Vi har också att

$$\bar{X} = \sigma \bar{Z} + \mu.$$

Om vi har täthetsfunktion $f_Z(\bar{z})$ får vi att

$$f_{\bar{X}}(\bar{x}) = \frac{1}{\sigma} f_Z\left(\frac{\bar{x} - \mu}{\sigma}\right).$$

Det följer att samplingfördelningarna för \bar{X} där X_1, \dots, X_n är oberoende och likafördelade från en läge-skalafamilj är också en läge-skalafamilj.

Exempel 4.3. Låt X_1, \dots, X_n vara oberoende och $\text{Exp}(\lambda)$ -fördelade; dvs. X_1, \dots, X_n har täthetsfunktion

$$f_X(x|\lambda) = \lambda e^{-\lambda x}.$$

Detta är en läge-skalafamilj med standardtäthetsfunktion

$$f_Z(z) = e^{-z}.$$

För Z_1, \dots, Z_n oberoende och $\text{Exp}(1)$ -fördelade såg vi i föreläsning 1 att

$$Y = \sum_{i=1}^n Z_i \sim \text{Gamma}(n, 1).$$

Så

$$f_Y(y) = \frac{1}{\Gamma(n)} y^{n-1} e^{-y}.$$

Det följer att $\bar{Z} = \frac{1}{n}Y$ har sannolikhetsfunktion

$$f_{\bar{Z}}(\bar{z}) = n f_Y(n\bar{z}) = \frac{n}{\Gamma(n)} (n\bar{z})^{n-1} e^{-n\bar{z}} = \frac{n^n}{\Gamma(n)} \bar{z}^{n-1} e^{-n\bar{z}}.$$

Det följer från den föregående diskussionen att

$$f_{\bar{X}}(\bar{x}) = \lambda f_{\bar{Z}}(\lambda \bar{x}) = \frac{n^n \lambda}{\Gamma(n)} (\lambda \bar{x})^{n-1} e^{-n\lambda \bar{x}} = \frac{(n\lambda)^n}{\Gamma(n)} \bar{x}^{n-1} e^{-(n\lambda)\bar{x}}.$$

Så vi får att $\bar{X} \sim \text{Gamma}(n, n\lambda)$.

4.1.1 Samplingfördelningar för normalfördelade stickprov

Med hjälp av satserna från de tidigare föreläsningarna kan vi hitta samplingfördelningarna för \bar{X} och S^2 när X_1, \dots, X_n är oberoende och $N(\mu, \sigma^2)$ -fördelade. Till exempel har vi att

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} X_1 + \dots + \frac{1}{n} X_n,$$

bara är en linjär transformation av $\mathbf{X} = (X_1, \dots, X_n) \sim N(\mu \mathbf{1}_n, \sigma^2 \mathbf{I}_n)$ där $\mathbf{1}_n = [1, 1, \dots, 1]^T \in \mathbb{R}^n$ och \mathbf{I}_n betecknar $(n \times n)$ -identitetsmatrisen. Sats 2.2 säger att för $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ och $A \in \mathbb{R}^{k \times n}$ en matris av full rang med $k \leq n$ och $\mathbf{b} \in \mathbb{R}^k$ har vi att $A\mathbf{X} + \mathbf{b} \sim N(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T)$. Om vi låter

$$A = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \in \mathbb{R}^{1 \times n}$$

har vi $\bar{X} \sim N(\mu, \sigma^2/n)$ eftersom

$$A\mu \mathbf{1}_n = \begin{bmatrix} \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \sum_{i=1}^n \frac{1}{n} \mu = \mu,$$

och

$$A\sigma \mathbf{I}_n A^T = \begin{bmatrix} \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix} = \frac{\sigma^2}{n}.$$

Så givet X_1, \dots, X_n oberoende och $N(\mu, \sigma^2)$ -fördelade har vi att $\bar{X} \sim N(\mu, \sigma^2/n)$. Vi kan också bestämma samplingfördelningen för S^2 med lite mer arbete. Först ska vi bevisa ett annat användbart faktum.

Sats 4.2. Låt X_1, \dots, X_n vara oberoende och $N(\mu, \sigma^2)$ -fördelade. Då är $\bar{X} \perp S^2$.

Bevis. Vi kan skriva S^2 som

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \\ &= \frac{1}{n-1} \left((X_1 - \bar{X})^2 + \sum_{i=2}^n (X_i - \bar{X})^2 \right), \\ &= \frac{1}{n-1} \left(\left(-\sum_{i=2}^n (X_i - \bar{X}) \right)^2 + \sum_{i=2}^n (X_i - \bar{X})^2 \right). \quad (\text{eftersom } \sum_{i=1}^n (X_i - \bar{X}) = 0) \end{aligned}$$

Så S^2 är en funktion av $X_2 - \bar{X}, \dots, X_n - \bar{X}$. Med hjälp av Sats 2.2 vet vi att $(\bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ är normalfördelad. Vi har också att

$$\begin{aligned} \text{Cov}[\bar{X}, X_i - \bar{X}] &= \text{Cov}[\bar{X}, X_i] - \text{Cov}[\bar{X}, \bar{X}], \\ &= \frac{1}{n} \sum_{j=1}^n \text{Cov}[X_j, X_i] - \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \text{Cov}[X_j, X_i], \\ &= \frac{1}{n} \sigma^2 - \frac{1}{n^2} n \sigma^2, \\ &= 0. \end{aligned}$$

Med hjälp av Sats 2.4 följer det att $\bar{X} \perp X_2 - \bar{X}, \dots, X_n - \bar{X}$ och därför är $\bar{X} \perp S^2$. □

För att bestämma samplingfördelningen av S^2 behöver vi några användbara fakta om χ^2 -fördelningen som vi ska använda utan bevis. Notisera dock att (b) kan visas med hjälp av observationen om en summa av oberoende gammafördelade stokastiska variabler som vi använde i Exempel 1.2 och (a) kan visas igenom att betrakta transformationen av variabler $Y = X^2$.

Lemma 4.1. Följande är fakta om χ^2 -fördelade stokastiska variabler:

(a) Om $Z \sim N(0, 1)$ då gäller $Z^2 \sim \chi^2(1)$.

(b) Om X_1, \dots, X_n är oberoende och $X_i \sim \chi^2(p_i)$ då gäller $X_1 + \dots + X_n \sim \chi^2(p_1 + \dots + p_n)$.

Nu kan vi bestämma samplingfördelningen för S^2 när X_1, \dots, X_n är oberoende och $N(\mu, \sigma^2)$ -fördelade.

Sats 4.3. Låt X_1, \dots, X_n vara oberoende och $N(\mu, \sigma^2)$ -fördelade. Då är

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1);$$

dvs. den stokastiska variabeln $Y = \frac{(n-1)S^2}{\sigma^2}$ har täthetsfunktion

$$f_Y(y) = \frac{1}{\Gamma(p/2)2^{p/2}} y^{\frac{p}{2}-1} e^{-\frac{y}{2}}.$$

Bevis. Anta att X_1, \dots, X_n är oberoende och $N(0, 1)$ -fördelade. I det som följer låt \bar{X}_k och S_k^2 beteckna stickprovsmedelvärde respektive stickprovsvariansen för de första k variabler i X_1, \dots, X_n . I uppgift 2 på övning 4 ska ni bevisa att

$$(n-1)S_n^2 = (n-2)S_{n-1}^2 + \left(\frac{n-1}{n}\right)(X_n - \bar{X}_{n-1})^2.$$

Vi ska bevisa satsen igenom induktion på n . Definiera $0 \times S_1^2 = 0$ och betrakta fallet $n = 2$. Vi har från vår rekursionsformel för $(n-1)S_n^2$ att

$$S_2^2 = \frac{1}{2}(X_2 - X_1)^2.$$

Med hjälp av Sats 2.2 ser vi att $(X_2 - X_1) \sim N(0, 2)$ och därmed är $\frac{X_2 - X_1}{\sqrt{2}} \sim N(0, 1)$. Från del (a) i Lemma 4.1 följer det att $S_2^2 \sim \chi^2(1)$.

Nu antar vi att $(n-1)S_n^2 \sim \chi^2(n-1)$ för $n = k$ och betraktar $n = k+1$. För $n = k+1$ får vi

$$kS_{k+1}^2 = (k-1)S_k^2 + \left(\frac{k}{k+1}\right)(X_{k+1} - \bar{X}_k)^2$$

från vår rekursionsformel. Det följer på samma sätt som i beviset av Sats 4.2 att $(k/(k+1))(X_{k+1} - \bar{X}_k)^2 \perp (k-1)S_k^2$. Vi har också att $(k/(k+1))(X_{k+1} - \bar{X}_k)^2 \sim \chi^2(1)$ eftersom $X_{k+1} - \bar{X}_k$ är normalfördelad med väntevärde 0 och varians

$$\text{Var}[X_{k+1} - \bar{X}_k] = \frac{k+1}{k}.$$

(Man kan bevisa detta faktum med hjälp av Sats 2.2.) Det följer att $(k/(k+1))(X_{k+1} - \bar{X}_k)^2 \sim \chi^2(1)$. Eftersom $(k-1)S_k^2 \sim \chi^2(k-1)$ enligt induktionshypotes får vi att $kS_{k+1}^2 \sim \chi^2(k)$ från Lemma 4.1 (b). Det följer att $(n-1)S_n^2 \sim \chi^2(n-1)$ när X_1, \dots, X_n är oberoende och $N(0, 1)$ -fördelade.

Det återstår att bevisa påståendet när X_1, \dots, X_n är oberoende och $N(\mu, \sigma^2)$ -fördelade. Först kom ihåg från Exempel 1.11 att normalfamiljen är en läge-skalfamilj med standardtäthetsfunktion given av $Z \sim N(0, 1)$. I synnerhet har vi att $X \sim N(\mu, \sigma)$ har täthetsfunktion

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right),$$

och $X_i = \sigma Z_i + \mu$ där Z_1, \dots, Z_n är oberoende och $N(0, 1)$ -fördelade. Låt $T(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2$. Då har vi

$$\begin{aligned} S^2 &= T(X_1, \dots, X_n), \\ &= T(\sigma Z_1 + \mu, \dots, \sigma Z_n + \mu), \\ &= \frac{1}{n-1} \sum_{i=1}^n \left(\sigma Z_i + \mu - \frac{1}{n} \sum_{j=1}^n (\sigma Z_j + \mu) \right)^2, \\ &= \frac{\sigma^2}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2. \end{aligned}$$

Låt $S_Z^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$. Det följer att $\frac{(n-1)S^2}{\sigma^2} = (n-1)S_Z^2$. Därför är

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

□

För att kvantifiera osäkerheten i \bar{X} när vi använder det som en skattning för μ kan vi använda samplingfördelningen

$$\bar{X} \sim N(\mu, \sigma^2/n),$$

eller motsvarande

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Detta fungerar bara när vi vet σ^2 . Om vi vet inte σ^2 kan vi skatta det som S^2 och skatta fördelningen $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ med kvoten

$$Y = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}},$$

där $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ och S^2/σ^2 är en skalad $\chi^2(n-1)$ -fördelad variabel. Eftersom $\bar{X} \perp S^2$ kan vi använda en transformation av stokastiska variabler för att beräkna sannolikhetsfunktionen för Y . Det är

$$f_Y(y) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)\sqrt{p\pi}} \frac{1}{\left(1 + \frac{y^2}{2}\right)^{(p+1)/2}}, \quad -\infty < y < \infty,$$

som kallas *t-fördelningen* med p frihetsgrader där $p = n - 1$. Generellt sett kan $p = 1, 2, \dots$ och vi säger att en stokastisk variabel Y med sådan täthetsfunktion $f_Y(y|p)$ är *t-fördelad med p frihetsgrader* och skriver $Y \sim t(p)$. Man kan visa att $E[Y] = 0$ om $p > 1$ och $\text{Var}[Y] = \frac{p}{p-2}$ om $p > 2$ för $Y \sim t(p)$.

4.2 Tillräckliga Statistikor

Vi såg i föregående avsnitt att \bar{X} och S^2 är bra skattningar för μ och σ^2 för en $N(\mu, \sigma^2)$ -fördelning och \bar{X} och S^2 berättar för oss allt vi behöver veta om fördelningen på ett sätt. Hur kan vi formulera denna relation för en godtycklig familj av fördelningar? Idéen är att studera fördelningen av $\mathbf{X} = (X_1, \dots, X_n)$ betingad på $T(\mathbf{X})$ (dvs, för $T(\mathbf{X})$ fixt). En statistika $T(\mathbf{X})$ kallas en *tillräcklig statistika* för parametern θ om fördelningen av \mathbf{X} givet $T(\mathbf{X})$ inte beror på θ .

Betrakta idén först för diskreta stokastiska variabler: Anta att $T(\mathbf{X})$ kan ta värdet t ; dvs. $P(T(\mathbf{X}) = t) > 0$. Vi är intresserade av fördelningen av \mathbf{X} given $T(\mathbf{X}) = t$; dvs.

$$P(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t).$$

Om $t \neq T(\mathbf{x})$ har vi att

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t) &= \frac{P(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t)}{P(T(\mathbf{X}) = t)}, \\ &= \frac{P(\mathbf{X} = \mathbf{x}, T(\mathbf{x}) = t)}{P(T(\mathbf{X}) = t)}, \\ &= 0. \end{aligned}$$

Alltså beror $P(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t)$ inte på θ . Om $t = T(\mathbf{x})$ har vi

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = t) &= P(\mathbf{X} = \mathbf{x} | T(\mathbf{X}) = T(\mathbf{x})), \\ &= \frac{P(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = T(\mathbf{x}))}{P(T(\mathbf{X}) = T(\mathbf{x}))}, \\ &= \frac{P(\mathbf{X} = \mathbf{x})}{P(T(\mathbf{X}) = T(\mathbf{x}))}, \end{aligned}$$

eftersom $\mathbf{X} = \mathbf{x}$ innebär att $T(\mathbf{X}) = T(\mathbf{x})$. Men detta kvot är lika med

$$\frac{P(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)},$$

där $P(\mathbf{x}|\theta)$ är sannolikhetsfunktionen av \mathbf{X} och $q(T(\mathbf{x})|\theta)$ är sannolikhetsfunktionen av $T(\mathbf{X})$. Så har vi visat att $T(\mathbf{X})$ är en tillräcklig statistika för θ om och endast om detta kvot beror inte på θ . Det samma resultatet stämmer för kontinuerliga fördelningar om vi ersätter sannolikhetsfunktioner för täthetsfunktioner.

Exempel 4.4. Låt X_1, \dots, X_n vara oberoende och $\text{Ber}(\theta)$ -fördelade. Vi skulle vilja visa att

$$T(\mathbf{X}) = X_1 + \dots + X_n$$

är en tillräcklig statistika för θ . Det vill säga att all nödvändig information i X_1, \dots, X_n för att skatta θ ges av antalet ettor i stickprovet (ordningen spelar ingen roll). Vi vet att

$$P(\mathbf{x}|\theta) = \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}.$$

Eftersom $T = T(\mathbf{X}) \sim \text{Bin}(\theta)$ vet vi också att

$$q(t|\theta) = \binom{n}{t} \theta^t (1-\theta)^{n-t}.$$

Eftersom $T(\mathbf{x}) = \sum_{i=1}^n x_i$ har vi att

$$\begin{aligned} \frac{P(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} &= \frac{\prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}}{\binom{n}{\sum_{i=1}^n x_i} \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}, \\ &= \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}{\binom{n}{\sum_{i=1}^n x_i} \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}}, \\ &= \frac{1}{\binom{n}{\sum_{i=1}^n x_i}}, \end{aligned}$$

vilken inte beror på θ . Så $T(\mathbf{X})$ är en tillräcklig statistika för θ .

Medan föregående metoden fungerar har den nackdelen att man måste först gissa $T(\mathbf{X})$ och därefter beräkna kvoten för att verifiera att det beror inte på θ . Följande sats ger oss en lösning till detta problem:

Sats 4.4 (Faktorisering sats). Låt $f(\mathbf{x}|\theta)$ vara täthetsfunktionen (eller sannolikhetsfunktionen) för $\mathbf{X} = (X_1, \dots, X_n)$ där X_1, \dots, X_n är oberoende och likafördelade, och låt $T(\mathbf{X})$ vara en statistika. Då är $T(\mathbf{X})$ en tillräcklig statistika med avseende på θ om och endast om det finns funktioner $g(t|\theta)$ och $h(\mathbf{x})$ så att

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

för alla \mathbf{x} och θ .

Bevis. Vi visar resultatet i det diskreta fallet. Anta först att $T(\mathbf{X})$ är en tillräcklig statistika. Vi har sett redan att

$$\frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)},$$

där $f(\mathbf{x}|\theta)$ är sannolikhetsfunktionen för \mathbf{X} och $q(T(\mathbf{x})|\theta)$ är sannolikhetsfunktionen för $T(\mathbf{X})$, inte beror på θ . Därför låter vi

$$h(\mathbf{x}) = \frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} \quad \text{och} \quad g(T(\mathbf{x})|\theta) = q(T(\mathbf{x})|\theta).$$

Anta nu att det finns funktioner $g(t|\theta)$ och $h(\mathbf{x})$ så att

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

för alla \mathbf{x} och θ . Vi ska visa att kvoten

$$\frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

inte beror på θ . Så vi måste beräkna $q(T(\mathbf{x})|\theta)$. Eftersom $q(T(\mathbf{x})|\theta)$ är sannolikhetsfunktionen för $T(\mathbf{X})$, vilken är en transformation av de diskreta stokastiska variablerna \mathbf{X} kan vi beräkna den som

$$\begin{aligned} q(T(\mathbf{x})|\theta) &= \sum_{\mathbf{y}: T(\mathbf{y})=T(\mathbf{x})} f(\mathbf{y}|\theta), \\ &= \sum_{\mathbf{y}: T(\mathbf{y})=T(\mathbf{x})} g(T(\mathbf{y})|\theta)h(\mathbf{y}), \\ &= \sum_{\mathbf{y}: T(\mathbf{y})=T(\mathbf{x})} g(T(\mathbf{x})|\theta)h(\mathbf{y}), \\ &= g(T(\mathbf{x})|\theta) \sum_{\mathbf{y}: T(\mathbf{y})=T(\mathbf{x})} h(\mathbf{y}). \end{aligned}$$

Det följer att

$$\frac{f(\mathbf{x}|\boldsymbol{\theta})}{q(T(\mathbf{x})|\boldsymbol{\theta})} = \frac{g(T(\mathbf{x})|\boldsymbol{\theta})h(\mathbf{x})}{g(T(\mathbf{x})|\boldsymbol{\theta}) \sum_{\mathbf{y}:T(\mathbf{y})=T(\mathbf{x})} h(\mathbf{y})} = \frac{h(\mathbf{x})}{\sum_{\mathbf{y}:T(\mathbf{y})=T(\mathbf{x})} h(\mathbf{y})},$$

som inte beror på $\boldsymbol{\theta}$. □

Exempel 4.5. Låt $\mathbf{X} = (X_1, \dots, X_n)$ där X_1, \dots, X_n är oberoende och $N(\mu, \sigma^2)$ -fördelade. Då är

$$f(\mathbf{x}|\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right).$$

Summan i exponenten kan skrivas som

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \mu)^2, \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + 2 \sum_{i=1}^n (x_i - \bar{x})(\bar{x} - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2, \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + 2(0) + \sum_{i=1}^n (\bar{x} - \mu)^2, \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2. \end{aligned}$$

Om σ^2 är känt får vi att

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) \exp(n(\bar{x} - \mu)^2), \\ &= h(\mathbf{x})g(\bar{x}|\mu), \end{aligned}$$

där

$$h(\mathbf{x}) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) \quad \text{och} \quad g(\bar{x}|\mu) = \exp(n(\bar{x} - \mu)^2).$$

Det följer att \bar{X} är en tillräcklig statistika för μ .

Om σ och μ är båda okända observerar vi att $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Så vi kan skriva

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp((n-1)s^2 + n(\bar{x} - \mu)^2), \\ &= h(\mathbf{x})g(\bar{x}, s^2|\mu, \sigma^2), \end{aligned}$$

där

$$h(\mathbf{x}) = 1 \quad \text{och} \quad g(\bar{x}, s^2|\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp((n-1)s^2 + n(\bar{x} - \mu)^2).$$

Så är (\bar{X}, S^2) en två-dimensionell tillräcklig statistika för (μ, σ^2) .

Exempel 4.6. Låt $\mathbf{X} = (X_1, \dots, X_n)$ där X_1, \dots, X_n är oberoende och $U(0, \theta)$ -fördelade för någon $\theta > 0$. Då är täthetsfunktionen

$$f_X(x|\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta, \\ 0 & \text{annat.} \end{cases}$$

Med hjälp av indikatorfunktionen för en mängd A

$$\mathbf{1}_A(x) = \begin{cases} 1 & x \in A, \\ 0 & x \notin A, \end{cases}$$

kan vi skriva täthetsfunktionen som

$$f_X(x|\theta) = \mathbf{1}_{[0, \theta]}(x).$$

Den simultana fördelningen är då

$$f(\mathbf{x}|\theta) = \frac{1}{\theta^n} \mathbf{1}_{[0,\theta]}(x_1) \cdots \mathbf{1}_{[0,\theta]}(x_n).$$

Eftersom $f(\mathbf{x}|\theta) \neq 0$ om och endast om $0 \leq x_1, \dots, x_n \leq \theta$ och $0 \leq x_1, \dots, x_n \leq \theta$ om och endast om $\min_{i \in [n]}(x_i) \geq 0$ och $\max_{i \in [n]}(x_i) \leq \theta$ kan vi skriva den simultana fördelningen som

$$\begin{aligned} f(\mathbf{x}|\theta) &= \frac{1}{\theta^n} \mathbf{1}_{[0,\infty)}(\min_{i \in [n]}(x_i)) \mathbf{1}_{(-\infty,\theta]}(\max_{i \in [n]}(x_i)), \\ &= h(\mathbf{x})g(T(\mathbf{x})|\theta), \end{aligned}$$

där

$$\begin{aligned} h(\mathbf{x}) &= \mathbf{1}_{[0,\infty)}(\min_{i \in [n]}(x_i)), \\ g(T(\mathbf{x})|\theta) &= \frac{1}{\theta^n} \mathbf{1}_{(-\infty,\theta]}(T(\mathbf{x})), \quad \text{och} \\ T(\mathbf{X}) &= \max_{i \in [n]}(X_i). \end{aligned}$$

Så det följer från Sats 4.4 att $T(\mathbf{X}) = \max_{i \in [n]}(X_i)$ är en tillräcklig statistika för θ . Resultatet stämmer eftersom resten av datan hjälper inte med skattning av θ när man har beräknat det största värdet i stickprovet.

Vi kan också få en formel för tillräckliga statistikor för en godtycklig exponentialfamilj.

Sats 4.5. Låt $\mathbf{X} = (X_1, \dots, X_n)$ där X_1, \dots, X_n är oberoende och likafördelade med täthetsfunktion (eller sannolikhetsfunktion) på formen

$$f_X(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) t_i(x) \right),$$

där $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ för $d \leq k$. Då är

$$\left(\sum_{j=1}^n t_1(X_j), \dots, \sum_{j=1}^n t_k(X_j) \right)$$

en k -dimensionell tillräcklig statistika för $\boldsymbol{\theta}$.

Bevis. Vi kan skriva den simultana täthetsfunktionen som

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) &= \prod_{j=1}^n h(x_j) c(\boldsymbol{\theta}) \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) t_i(x_j) \right), \\ &= \left(\prod_{j=1}^n h(x_j) \right) c(\boldsymbol{\theta})^n \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) \sum_{j=1}^n t_i(x_j) \right), \\ &= h(\mathbf{x})g(T(\mathbf{x})|\boldsymbol{\theta}), \end{aligned}$$

där

$$\begin{aligned} h(\mathbf{x}) &= \prod_{j=1}^n h(x_j), \quad \text{och,} \\ g(T(\mathbf{x})|\boldsymbol{\theta}) &= c(\boldsymbol{\theta})^n \exp \left(\sum_{i=1}^k w_i(\boldsymbol{\theta}) \sum_{j=1}^n t_i(x_j) \right). \end{aligned}$$

Påståendet följer från Sats 4.4. □

4.3 Punktskattning

Nu när vi vet några metoder för att hitta bra statistikor är frågan, ”Hur kan man använda sådana statistikor för att skatta parametrarna?” Anta, exempelvis, att x_1, \dots, x_n är ett stickprov där X_1, \dots, X_n är oberoende och $N(\mu, 4)$ -fördelade. Vi skulle vilja använda våra statistikor för att skatta μ . Vi såg i Exempel 4.5 att \bar{X} är en tillräcklig statistika för μ ; dvs. all information som man kunde använda för att skatta μ ingår i \bar{X} . Vi såg också att $E[\bar{X}] = \mu$. Dessa fakta föreslår att en bra skattning för μ är $\tilde{\mu} = \bar{x}$. Detta val är ganska enkelt men är helt baserat på vad vi vet om μ igenom datan x_1, \dots, x_n .

Generellt sett kallas denna process för att skatta parametrarna $\theta_1, \dots, \theta_d$ i en fördelning $f_X(x|\theta_1, \dots, \theta_d)$ från data för en *punktskattning* och det görs med en funktion

$$W : \mathbb{R}^n \longrightarrow \mathbb{R}^d$$

där $W(\mathbf{X}) = W(X_1, \dots, X_n)$. Funktionen W eller $W(\mathbf{X})$ kallas en *skattning* eller *punktskattning* och det observerade värdet $W(\mathbf{x}) = W(x_1, \dots, x_n)$ kallas det samma. Vår första metod för att hitta en punktskattning generaliserar vårt lilla exempel med normalfördelningen.

4.3.1 Momentmetoden

En av de enklaste metoderna för att bestämma en punktskattning är följande: Anta att vi har $\mathbf{X} = (X_1, \dots, X_n)$ där X_1, \dots, X_n är oberoende och likafördelade med sannolikhetsfunktion (eller täthetsfunktion) $f_X(x|\boldsymbol{\theta})$ där $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$. Vi kan beräkna de första d *stickprovsmomenten*:

$$\begin{aligned} m_1(\mathbf{X}) &= \frac{1}{n} \sum_{i=1}^n X_i, \\ m_2(\mathbf{X}) &= \frac{1}{n} \sum_{i=1}^n X_i^2, \\ &\vdots \\ m_d(\mathbf{X}) &= \frac{1}{n} \sum_{i=1}^n X_i^d. \end{aligned}$$

Vi kan också beräkna formeler för de första d *populationsmomenten*:

$$E[X|\boldsymbol{\theta}], E[X^2|\boldsymbol{\theta}], \dots, E[X^d|\boldsymbol{\theta}].$$

Då får vi ett system av ekvationer

$$\begin{aligned} m_1(\mathbf{X}) &= E[X|\boldsymbol{\theta}], \\ m_2(\mathbf{X}) &= E[X^2|\boldsymbol{\theta}], \\ &\vdots \\ m_k(\mathbf{X}) &= E[X^k|\boldsymbol{\theta}]. \end{aligned}$$

som vi kan lösa för att få en *momentmetodens punktskattning*:

$$\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \dots, \tilde{\theta}_d).$$

Exempel 4.7. Låt x_1, \dots, x_n vara ett stickprov där X_1, \dots, X_n är oberoende och $N(\mu, \sigma^2)$ -fördelade. Vi skulle vilja bestämma en momentmetodens punktskattning för (μ, σ^2) . Det första momentet av $X \sim N(\mu, \sigma^2)$ är $E[X] = \mu$. För att bestämma det andra momentet använder vi att $\text{Var}[X] = \sigma^2$ och

$$\text{Var}[X] = E[X^2] - E[X]^2.$$

Det följer att

$$\begin{aligned} E[X^2] &= \text{Var}[X] + E[X]^2, \\ &= \sigma^2 + \mu^2. \end{aligned}$$

De första två stickprovsmomenten är

$$m_1(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad \text{och,}$$
$$m_2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Så löser vi systemet

$$\bar{X} = \mu,$$
$$\frac{1}{n} \sum_{i=1}^n X_i^2 = \sigma^2 + \mu^2.$$

Den första ekvationen ger oss att $\tilde{\mu} = \bar{x}$. Så ger oss den andra ekvationen att

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2, \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x}^2 + \bar{x}^2, \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{2\bar{x}}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \bar{x}^2, \\ &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2), \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \\ &= \frac{n-1}{n} s^2. \end{aligned}$$

Så får vi punktskattningen $(\tilde{\mu}, \tilde{\sigma}^2) = (\bar{x}, (n-1)s^2/n)$ för (μ, σ^2) . Punktskattningen för μ är bra eftersom $E[\tilde{\mu}] = \mu$. Å andra sidan har vi att

$$E[\tilde{\sigma}^2] = \frac{n-1}{n} \sigma^2.$$

Så skattningen för σ^2 kommer att underskatta det rätta värdet för små n . Vi kommer att diskutera detta problem i den kommande föreläsningen.

Exempel 4.8. Låt x_1, \dots, x_n vara ett stickprov där X_1, \dots, X_n är oberoende och $\text{Bin}(\theta, p)$ -fördelade där både θ och p okända. Vi skulle vilja bestämma en momentmetodens punktskattning för (θ, p) . Det första momentet av $X \sim \text{Bin}(\theta, p)$ är $E[X] = \theta p$. För att bestämma det andra momentet använder vi att $\text{Var}[X] = \theta p(1-p)$ och $\text{Var}[X] = E[X^2] - E[X]^2$. Då får vi att

$$\begin{aligned} E[X^2] &= \text{Var}[X] + E[X]^2, \\ &= \theta p(1-p) + \theta^2 p^2. \end{aligned}$$

De första två stickprovsmomenten är

$$m_1(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad \text{och}$$
$$m_2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Då löser vi systemet av ekvationer:

$$\bar{x} = \theta p,$$
$$m_2(\mathbf{x}) = \theta p(1-p) + \theta^2 p^2.$$

Den första ekvationen ger

$$p = \frac{\bar{x}}{\theta}.$$

Med hjälp av denna formel och den andra ekvationen får vi att

$$\tilde{\theta} = \frac{\bar{x}^2}{\bar{x} - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Formeln för $\tilde{\theta}$ ger oss att

$$\tilde{p} = \frac{\bar{x}}{\tilde{\theta}} = \frac{\bar{x} - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}{\bar{x}}.$$

Denna punktskattning kan vara opålitlig när n är liten eftersom vi kan ha

$$\bar{x} < \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

vilket ger $\tilde{\theta} < 0$. Men det är en bra start eftersom den är lätt att beräkna. I den nästa föreläsningen ska vi se andra metoder för punktskattning.