# Teoretisk statistik

Gunnar Englund
Matematisk statistik
KTH

Vt 2005

## 1 Inledning

Vi skall kortfattat behandla aspekter av teoretisk statistik där framför allt begreppet ”uttömmande” (ibland kallad ”tillräcklig”) stickprovsvariabel är viktigt. En sådan innebär en reduktion av data utan att man för den sakens skull tappar någon relevant information rörande parametern. Vi kommer att kunna visa en enkel undre gräns för variansen för väntevärdesriktiga skattningar som man dessutom kan uppnå i vissa sammanhang. Detta innebär alltså att man kan visa att vissa skattningar är de effektivaste av alla tänkbara skattningar.

## 2 Uttömmande stickprovsvariabler

En punktskattning $t = t(x_1, x_2, \ldots, x_n) = t(\boldsymbol{x})$ av en parameter $\theta$ innebär en reduktion av informationen i mätdata $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Om vi studerar de $\boldsymbol{x}$ sådana att $t(\boldsymbol{x}) = t$ är detta en delmängd av $R^n$. På motsvarande sätt är för olika $t$ händelserna

$$A_t = \{\omega \in \Omega : t(X_1, X_2, \ldots, X_n) = t\}$$

delmängder av $\Omega$. Dessa utgör en s k partition (uppdelning eller sönderläggning) $\Pi$ av $\Omega$, d v s $\Omega = \bigcup_t A_t$ och $\Pi = \{A_t,\ t \in R\}$. Olika stickprovsvariabler ger olika partitioner av $\Omega$. En partition $\Pi_1$ är en reduktion av en partition $\Pi_2$ om mängderna i $\Pi_1$ alla utgörs av unioner av mängderna i $\Pi_2$. Partitionen $\Pi_1$ innebär då en ”hopklumpning” av partitionen $\Pi_2$. Vi kommer att vara intresserade av att finna den mest reducerade partitionen – mer om detta i avsnittet om minimala uttömmande stickprovsvariabler.
<u>Exempel</u>: Vi kan definiera uppdelningarna av $R^n$ t ex genom att dela upp efter värdet på $x_1^2 + x_2^2 + \cdots + x_n^2$. Detta innebär att att dessa blir koncentriska sfärer i $R^n$. Skulle vi dessutom dela upp efter värdet på $x_1 + x_2 + \cdots + x_n$ svarar det mot plan. Kombinationen av dessa två skulle ge cirklar i $R^n$ med olika radier och i olika orienteringar. Dessa uppdelningar av $R^n$ i form av cirklar kan alltså ses som genererade av i detta fall 2 stickprovsvariabler. I

$\Omega$ svarar denna uppdelning av $R^n$ mot händelserna att variablerna hamnar i motsvarande cirklar och utgör alltså en partition av $\Omega$. $\square$

Exempel: Antag att vi har tre oberoende upprepningar av slantsingling som resulterar i värdet 0 eller 1 med sannolikhet $1-p$ respektive $p$. Vi har alltså data $x_1, x_2, x_3$ som alla är 0 eller 1. Vi har nu stickprovsvariabeln $t(x_1, x_2, x_3) = x_1 + x_2 + x_3 = k$ där alltså $k$ står för antalet 1:or. $X_1 + X_2 + X_3$ är som bekant $\mathrm{Bin}(3, p)$. Vi får

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3} = p^k(1-p)^{3-k}.$$

För fixt $k$ utgör $A_k = \{\omega \in \Omega : X_1 + X_2 + X_3 = k\}$ en del av $\Omega$. För olika värden på $k$ erhålls en partition av $\Omega$. Vi låter $T = X_1 + X_2 + X_3$ som alltså $k$ är ett utfall av. Vi erhåller

$$\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid T = k) = \\
&\frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_1 + X_2 + X_3 = k)}{P(X_1 + X_2 + X_3 = k)} = \\
&\begin{cases} 0, & \text{om } x_1 + x_2 + x_3 \neq k \\ \dfrac{p^k(1-p)^{3-k}}{\binom{3}{k}p^k(1-p)^{3-k}} = \dfrac{1}{\binom{3}{k}}, & \text{om } x_1 + x_2 + x_3 = k. \end{cases}
\end{aligned}$$

Notera att fördelningen för $X_1, X_2, X_3$ givet värdet på $X_1 + X_2 + X_3$ ej beror av parametern $p$. Detta betyder att givet kunskap om värdet på $X_1 + X_2 + X_3$, d v s antalet 1:or, saknar den ytterligare kunskapen om i vilken ordning 0:orna och 1:orna kommit relevans när det gäller $p$. $\square$

Vi är nu mogna för följande definition.

Definition: En stickprovsvariabel $T = t(X_1, X_2, \ldots, X_n)$ är uttömmande för en familj fördelningar om betingade fördelningen för $X_1, X_2, \ldots, X_n$ givet värdet på $T$ är den samma för alla medlemmar i familjen. $\square$

I allmänhet är familjen bestämd av en parameter som $p$ i exemplet ovan. Att $T = t_0$ innebär att $X_1, \ldots, X_n$ ligger i partitionen $t(x_1, \ldots, x_n) = t_0$. Att $T$ är uttömmande innebär att fördelningen för $X_1, \ldots, X_n$ på partitionen ej beror av parametern och alltså inte ger någon information om parametern. Man kan se datagenereringen som utförd i två steg där man först får värdet på den uttömmande stickprovsvariabeln. Denna lottning påverkas i hög grad av värdet på parametern. Man vet då vilken partition som data hamnat i. I steg två väljs en punkt på partitionen och denna lottning beror ej av parametern. I exemplet med binomialfördelade data var det t o m likformig fördelning på partitionen, men så snällt behöver det inte alltid bli.

Följande sats ger ett enkelt kriterium på när en stickprovsvariabel $T$ är uttömmande.

**Sats 1** En stickprovsvariabel $T = t(X_1, X_2, \ldots, X_n)$ är uttömmande för parametern $\theta$ om och endast om tätheten (eller sannolikhetsfunktionen) för

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ kan skrivas

$$f_{\boldsymbol{X}}(\boldsymbol{x}, \theta) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta) = g(t(\boldsymbol{x}), \theta) h(\boldsymbol{x}),$$

d v s som en funktion av $\theta$ och $t(x_1, x_2, \ldots, x_n)$ och en funktion som bara beror av $x_1, x_2, \ldots, x_n$. $\square$

Bevis (diskreta fallet):
($\Leftarrow$:) Antag att $P(\boldsymbol{X} = \boldsymbol{x}) = g(t(\boldsymbol{x}), \theta) h(\boldsymbol{x})$. Vi får fördelningen för $T$ genom att summera över de $\boldsymbol{x}$ där $t(\boldsymbol{x})$ är konstant. Vi får

$$P(T = t_0) = \sum_{t(\boldsymbol{x}) = t_0} P(\boldsymbol{X} = \boldsymbol{x}) = \sum_{t(\boldsymbol{x}) = t_0} g(t(\boldsymbol{x}), \theta) h(\boldsymbol{x}) = g(t_0, \theta) \sum_{t(\boldsymbol{x}) = t_0} h(\boldsymbol{x}).$$

Den betingade fördelningen för $\boldsymbol{X}$ givet $T = t_0$ blir alltså

$$P(\boldsymbol{X} = \boldsymbol{x} | T = t_0) = \frac{P(\boldsymbol{X} = \boldsymbol{x}, T = t_0)}{P(T = t_0)} = \begin{cases} 0, & \text{om } t(\boldsymbol{x}) \neq t_0 \\ \dfrac{h(\boldsymbol{x})}{\sum_{t(\boldsymbol{y}) = t_0} h(\boldsymbol{y})}, & \text{om } t(\boldsymbol{x}) = t_0. \end{cases}$$

Notera att detta ej beror av parametern $\theta$. Alltså är $T$ uttömmande. Denna betingade fördelning är alltså fördelningen över partitionen givet värdet av den uttömmande stickprovsvariabeln. Skulle $h(\boldsymbol{x})$ vara konstant svarar den mot likformig fördelning över partitionen. I allmänhet beror dock $h$ av $\boldsymbol{x}$.

($\Rightarrow$:) Antag att $T$ är uttömmande. Då gäller alltså att $P(\boldsymbol{X} = \boldsymbol{x} \mid T = t_0)$ ej skall bero av $\theta$. Vi har alltså $P(\boldsymbol{X} = \boldsymbol{x} \mid T = t_0) = c(\boldsymbol{x}, t_0)$. Vi får om $t(\boldsymbol{x}) = t_0$

$$\begin{aligned} &P(\boldsymbol{X} = \boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x}, t(\boldsymbol{X}) = t_0) = \\ &P(\boldsymbol{X} = \boldsymbol{x} \mid t(\boldsymbol{X}) = t_0) P(t(\boldsymbol{X}) = t_0) = c(\boldsymbol{x}, t_0) P(t(\boldsymbol{X}) = t(\boldsymbol{x})), \end{aligned}$$

som utgör den önskade faktoriseringen eftersom $P(t(\boldsymbol{X}) = t(\boldsymbol{x}))$ bara beror av $\boldsymbol{x}$ genom värdet på $t(\boldsymbol{x})$. $\square$

I binomialexemplet ovan gäller att

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = p^{x_1 + x_2 + x_3} (1 - p)^{3 - x_1 - x_2 - x_3} = p^k (1 - p)^{3 - k},$$

som t o m bara beror av $x_1 + x_2 + x_3$. Alltså är $T = X_1 + X_2 + X_3$ uttömmande.

Exempel:Poisson-fördelade data
Låt $X_1, X_2, \ldots, X_n$ vara oberoende Po($\mu$)-fördelade. Vi har då om $T(\boldsymbol{x}) = x_1 + x_2 + \cdots + x_n = k$

$$P(\boldsymbol{X} = \boldsymbol{x}) = \frac{\mu^{x_1 + x_2 + \cdots + x_n} e^{-n\mu}}{x_1! x_2! \ldots x_n!} = \mu^k e^{-n\mu} \cdot \frac{1}{x_1! x_2! \ldots x_n!}$$

så detta faktoriseras på rätt sätt med $h(\boldsymbol{x}) = 1/(x_1! \ldots x_n!)$ och $g(T(\boldsymbol{x}), \mu) = \mu^{T(\boldsymbol{x})} e^{-n\mu}$. Alltså är $T(\boldsymbol{X}) = X_1 + X_2 + \cdots + X_n$ uttömmande. Man behöver inte hålla reda på de enskilda observationerna utan bara deras summa. $\square$

I satsen ovan ser man att det finns ett antal uttömmande stickprovsvariabler, t ex är $\boldsymbol{X}$ faktiskt uttömmande. Om data är oberoende likafördelade är de s k ordningsvariablerna $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ som är $X_1, X_2, \ldots, X_n$ ordnade i storleksordning uttömmande. Mer intressant är det naturligtvis om den uttömmande stickprovsvariabeln innebär en väsentlig reduktion av data. Vi återkommer senare till problemet att finna en minimal uttömmande stickprovsvariabel, dvs en som inte ytterligare kan reduceras.

# 3 Likelihood och information

Att $t(\boldsymbol{X})$ är uttömmande för $\theta$ innebär att likelihoodfunktionen kan skrivas

$$L(\theta) = f(\boldsymbol{x}; \theta) = g(t(\boldsymbol{x}), \theta)h(\boldsymbol{x})$$

vilket ju innebär att ML-skattningen bara beror av $\boldsymbol{x}$ genom värdet på $t(\boldsymbol{x})$. Detta syns tydligast om vi logaritmerar och erhåller

$$\ln L(\theta) = \ln(g(t(\boldsymbol{x}), \theta)) + \ln(h(\boldsymbol{x})),$$

där ju sista termen ej beror av $\theta$ och alltså försvinner då man deriverar med avseende på $\theta$.

Vi inför den stokastiska variabeln

$$W = \frac{\partial}{\partial\theta}\ln(L(\theta, \boldsymbol{X})) = \frac{L'(\theta)}{L(\theta)} = \frac{f'(\boldsymbol{X}, \theta)}{f(\boldsymbol{X}, \theta)},$$

där $'$ står för derivering med avseende på $\theta$. Variabeln $W$ kallas ibland score.

Exempel:Poisson-fördelade data.
Med $n$ oberoende Po($\mu$)-fördelade data erhålls

$$L(\mu) = \mu^{x_1+\cdots+x_n}e^{-n\mu}/(x_1!\ldots x_n!)$$

som ger

$$\ln(L(\mu)) = (x_1 + x_2 + \cdots + x_n) \cdot \ln(\mu) - n\mu - \ln(x_1!\ldots x_n!)$$

som deriverat med avseende på $\mu$ ger

$$W = \frac{X_1 + X_2 + \cdots + X_n}{\mu} - n.$$

Vi noterar att $E(W) = 0$ och detta är ingen tillfällighet som framgår av nästa sats. □

Sats: Scoren $W = \frac{\partial}{\partial\theta}\ln L(\theta; \boldsymbol{X}) = f'(\boldsymbol{X}; \theta)/f(\boldsymbol{X}; \theta)$ har väntevärde 0, d v s $E(W) = 0$ oavsett värdet på $\theta$.

Bevis: Vi får

$$E(W) = E(\frac{f'(\boldsymbol{X}, \theta)}{f(\boldsymbol{X}, \theta)}) = \int_{R^n}\frac{f'(\boldsymbol{x}, \theta)}{f(\boldsymbol{x}, \theta)}f(\boldsymbol{x}, \theta)d\boldsymbol{x} = \frac{d}{d\theta}\int_{R^n}f(\boldsymbol{x}, \theta)d\boldsymbol{x} = \frac{d}{d\theta}(1) = 0,$$

där vi antar att vi kan flytta ut deriveringen m a p $\theta$ ur integralen. $\square$

Variansen för $W$ spelar stor roll i fortsättningen och vi kommer att kalla den informationen.

Definition: Informationen från ett stickprov $\boldsymbol{X}$ rörande parametern $\theta$ är $I_{\boldsymbol{X}}(\theta) = V(W)$. $\square$

För Poisson-fördelade data där $W = (X_1 + X_2 + \cdots + X_n)/\mu - n$ enligt ovan erhålls informationen

$$I_{\boldsymbol{X}}(\mu) = V(W) = \frac{1}{\mu^2}(V(X_1) + V(X_2) + \cdots + V(X_n)) = \frac{n\mu}{\mu^2} = \frac{n}{\mu}.$$

Notera att informationen, trots beteckningen inte är stokastisk. Man kan jämföra med beteckningen $f_{\boldsymbol{X}}$ för tätheten för $\boldsymbol{X}$.

Eftersom $E(W) = 0$ är $V(W) = E(W^2)$ och vi erhåller

$$I_{\boldsymbol{X}}(\theta) = E((\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X},\theta))^2) = E\left((\frac{f'(\boldsymbol{X};\theta)}{f(\boldsymbol{X};\theta)})^2\right).$$

Man noterar att om vi har två oberoende datamängder $\boldsymbol{X}$ och $\boldsymbol{Y}$ och lägger ihop dessa till en enda datamängd $\boldsymbol{X}, \boldsymbol{Y}$ så erhålls

$$W_{\boldsymbol{X},\boldsymbol{Y}} = \frac{\partial}{\partial\theta}\ln(L(\theta;\boldsymbol{X},\boldsymbol{Y})) = \frac{\partial}{\partial\theta}\ln(L(\theta;\boldsymbol{X})L(\theta;\boldsymbol{Y})) = W_{\boldsymbol{X}} + W_{\boldsymbol{Y}}$$

och alltså att $I_{\boldsymbol{X},\boldsymbol{Y}} = I_{\boldsymbol{X}} + I_{\boldsymbol{Y}}$ eftersom variansen för en summa av oberoende variabler är summan av varianserna.

Som specialfall kan vi ta situationen med $n$ oberoende likafördelade observationer
$X_1, X_2, \ldots, X_n$ där vi alltså får

$$I_{X_1,X_2,\ldots,X_n} = I_{X_1} + I_{X_2} + \cdots + I_{X_n} = nI_{X_1}.$$

I Poisson-exemplet är scoren för första observationen $W = X_1/\mu - 1$ och alltså informationen $V(W) = V(X_1/\mu - 1) = V(X_1)/\mu^2 = \mu/\mu^2 = 1/\mu$. För ett stickprov av $n$ oberoende observationer erhålls alltså informationen $n/\mu$.

Om vi har en uttömmande stickprovsvariabel $T(\boldsymbol{X})$ så har $T(\boldsymbol{X})$ samma information om parametern som $\boldsymbol{X}$ själv.

Sats: Om $T(\boldsymbol{X})$ är uttömmande för $\theta$ gäller att $I_{T(\boldsymbol{X})} = I_{\boldsymbol{X}}$. $\square$

Bevis: Eftersom $T(\boldsymbol{X})$ är uttömmande kan likelihooden skrivas

$$f(\boldsymbol{x},\theta) = g(t(\boldsymbol{x}),\theta)h(\boldsymbol{x})$$

så vi erhåller $f(\boldsymbol{X},\theta) = g(t(\boldsymbol{X}),\theta)h(\boldsymbol{X})$ som ger

$$\ln(f(\boldsymbol{X},\theta)) = \ln(g(t(\boldsymbol{X}),\theta)) + \ln(h(\boldsymbol{X})).$$

Deriveras detta med avseende på $\theta$ försvinner sista termen som ju ej beror av $\theta$ och vi erhåller scoren

$$W = \frac{\partial}{\partial \theta} \ln(f(\boldsymbol{X}, \theta)) = \frac{\partial}{\partial \theta} \ln(g(t(\boldsymbol{X}), \theta)),$$

som medför påståendet. $\square$

Informationen hänger intimt samman med variansen för skattningar enligt följande. Betrakta scoren $W$ för observationerna $\boldsymbol{X}$ relativt parametern $\theta$. Vi hade ju $E(W) = 0$ och vi studerar nu kovariansen mellan $T = t(\boldsymbol{X})$ och $W$. Vi erhåller

$$C(T, W) = E(TW) - E(T)E(W) = E(TW) = E\left(T \frac{\partial}{\partial \theta} \ln(f(\boldsymbol{X}, \theta))\right) =$$

$$\int_{R^n} t(\boldsymbol{x}) \frac{1}{f(\boldsymbol{x}, \theta)} (\frac{\partial}{\partial \theta} f(\boldsymbol{x}, \theta)) f(\boldsymbol{x}, \theta) d\boldsymbol{x} = \frac{d}{d\theta} \int_{R^n} t(\boldsymbol{x}) f(\boldsymbol{x}, \theta) d\boldsymbol{x} = \frac{d}{d\theta} E(T).$$

Om nu $T$ är väntevärdesriktig så gäller $E(T) = \theta$ och detta ger följande mycket intressanta resultat: $C(W, T) = 1$. Vi vet ju att korrelationskoeficienten ligger mellan $-1$ och 1 och detta ger

$$\varrho(W, T) = \frac{C(W, T)}{\sqrt{V(W)V(T)}} = \frac{1}{\sqrt{V(W)V(T)}}$$

dvs $V(T) \geq 1/V(W)$ ty $|\varrho(W, T)| \leq 1$. Detta ger den fundamentala Cramér-Raos sats.

<u>Sats:</u>(Cramér-Raos sats)
Om $T = t(\boldsymbol{X})$ är väntevärdesriktig gäller $V(T) \geq 1/V(W) = 1/I(\theta)$. $\square$

Vi har alltså knutit ihop informationen med en undre gräns för variansen för en skattning. Om man nu har hittat en väntevärdesriktig skattning $T$ av $\theta$ sådan att $V(T) = 1/V(W)$ kan man dra slutsatsen att det inte finns någon annan väntevärdesriktig skattning som är effektivare, dvs har mindre varians.

<u>Exempel:</u> (forts.) Med Poisson-fördelade data fick vi $V(W) = n/\mu$ och eftersom $T = \bar{X}$ är en väntevärdesriktig skattning med variansen

$$V(T) = V(\frac{1}{n}(X_1 + X_2 + \ldots, X_n)) = \frac{1}{n^2} n V(X_1) = \frac{\mu}{n} = \frac{1}{V(W)},$$

vilket innebär att vi uppnått undre gränsen i satsen ovan. Alltså finns ingen annan effektivare väntevärdesriktig skattning av $\mu$. $\square$

# 4 Exponentiella familjer

Den exponentiella (en-parametriga) familjen fördelningar definieras av att

$$f(x;\theta) = B(\theta)h(x)\exp(Q(\theta)R(x))$$

för lämpliga val av funktioner. Vi tolkar $f(x;\theta)$ antingen som en täthet eller som sannolikhetsfunktion. T ex kan $\mathrm{Bin}(n,\theta)$-fördelningens sannolikhetsfunktion skrivas

$$p(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = (1-\theta)^n\binom{n}{x}\exp(x\ln(\theta/(1-\theta))),\ x = 0, 1, \ldots, n$$

dvs i denna form med $B(\theta) = (1-\theta)^n$, $h(x) = \binom{n}{x}$, $Q(\theta) = \ln(\theta/(1-\theta))$ och $R(x) = x$. Även Poisson-fördelningen, för-första-gången-fördelningen och exponentialfördelningen är exempel på en-parametriga exponentiella familjer.

Om vi granskar Cramér-Raos sats ser vi att likhet uppstår då korrelationen mellan $T$ och $W$ är 1. Detta innebär att $W$ kan skrivas som ett linjärt uttryck i $T$, dvs att

$$W = \frac{\partial}{\partial\theta}\ln f(\boldsymbol{X};\theta) = c(\theta)T + d(\theta),$$

där vi kallat koefficienterna $c$ och $d$. Dessa får bero av $\theta$ men ej av $\boldsymbol{X}$. Integreras detta uttryck erhålls

$$\ln(f(\boldsymbol{X};\theta)) = C(\theta)T + D(\theta) + K(\boldsymbol{X})$$

där $C'(\theta) = c(\theta)$ och $D'(\theta) = d(\theta)$. Detta innebär att $f(\boldsymbol{x};\theta)$ är en medlem av en exponentiell familj eftersom den kan skrivas

$$f(\boldsymbol{x};\theta) = \exp(D(\theta))\exp(K(\boldsymbol{x}))\exp(C(\theta)t(\boldsymbol{x})),$$

där alltså $B(\theta) = \exp(D(\theta))$, $h(x) = \exp(K(\boldsymbol{x}))$, $Q(\theta) = C(\theta)$ samt $R(\boldsymbol{x}) = t(\boldsymbol{x})$.

Å andra sidan gäller att om $f(x;\theta)$ tillhör en exponentiell familj

$$f(x;\theta) = B(\theta)h(x)\exp(Q(\theta)R(x))$$

så är $\sum_1^n R(X_i)$ en uttömmande stickprovsvariabel. Alltså uppnås den undre gränsen i Cramér-Raos sats och i klassen av väntevärdesriktiga skattningar har vi hittat en med minimal varians, dvs som är effektivast.

Det finns även fler-parametriga exponentiella familjer som har tätheter (sannolikhetsfunktioner) av typen

$$f(x;\theta) = B(\theta)h(x)\exp(Q_1(\theta)R_1(x) + \cdots + Q_k(\theta)R_k(x)),$$

där $\theta$ är $k$-dimensionell. Normalfördelningen och $\Gamma$-fördelningen är av denna typ för $k = 2$. Vidare är $(R_1(\boldsymbol{X}), R_2(\boldsymbol{X}), \ldots, R_k(\boldsymbol{X}))$ en uppsättning uttömmande stickprovsvariabler. Man kan visa att om man har $n$ observationer och det finns en $k$-dimensionell uttömmande stickprovsvariabel där $k < n$ så kommer data från en $k$-parametrig exponentiell familj. De exponentiella familjerna dyker alltså upp då man kan göra en reduktion av data till en lägre dimension än data själv.

# 5 Konstruktion av effektivare skattningar

Om vi funderar på att använda $U = u(\boldsymbol{X})$ som skattning av $\theta$ och vi har tillgång till en uttömmande stickprovsvariabel $T = t(\boldsymbol{X})$ så kan vi ur $U$ konstruera en ny skattning med mindre varians än $U$. Låt nämligen $g(t) = E(U|T = t)$, dvs betingade väntevärdet givet $T = t$. Om vi nu sätter $Y = g(T)$ (som brukar betecknas $Y = E(U|T)$) så gäller enligt lagen om total förväntan att $E(Y) = E(E(U|T)) = E(U)$. Alltså är $Y$ väntevärdesriktig om $U$ är det. Vidare gäller enligt sats 5.16 i läroboken att

$$V(U) = E(V(U|T)) + V(E(U|T)) = E(V(U|T)) + V(Y).$$

Eftersom $V(U|T) \geq 0$ så gäller alltså $V(U) \geq V(Y)$. Slutsatsen är alltså att om en skattning inte skulle vara en funktion av en uttömmande stickprovsvariabel så kan man "förbättra" den genom att betinga med avseende på en uttömmande stickprovvariabel. Vi har alltså följande sats.

Sats: (Rao-Blackwells sats).

Om $U$ är en skattning och $T$ en uttömmande stickprovsvariabel och $Y = E(U|T)$ så gäller att $E(Y) = E(U)$ och att $V(U) \geq V(Y)$. □

Man får t ex för väntevärdesriktiga förslag $U$ med denna metod (Rao-Blackwellization) en väntevärdesriktig skattning som är effektivare än $U$. F ö kan man notera att en sådan skattning är en funktion av en uttömmande stickprovsvariabel.

I nästa avsnitt kommer vi att ta fram s k minimala uttömmande stickprovsvariabler. Dessa innebär en maximal reduktion av data utan att ha förlorat någon information rörande parametern. Om man har en minimal uttömmande stickprovsvariabel $T$ så definierar den en partition som är maximalt "grov". Partitionen av en annan uttömmande (men inte minimal) stickprovsvariabel $U$ är alltså sådan att medlemmarna i dess partition helt ligger i enskilda medlemmar av partitionerna för den minimalt uttömmande. Olikheten i Rao-Blackwells sats blir en likhet om och endast om $P(V(U|T) = 0) = 1$ som ju innebär att $U$ är en funktion av $T$. I så fall har ju $U|T = t$ en enpunktsfördelning. Omvänt gäller att om $U$ givet $T$ har betingade variansen 0 har den med sannolikhet 1 värdet $E(U|T = t) = g(t)$. Detta betyder att med ett $\boldsymbol{X} = \boldsymbol{x}$ med $t(\boldsymbol{x}) = t$ är värdet av $U$ just $g(t) = g(t(\boldsymbol{x}))$, dvs att $U$ är en funktion av $T$. Om Rao-Blackwellization inte ger en minskning av variansen så är redan $U$ en funktion av $T$.

# 6 Minimala uttömmande stickprovsvariabler

Man kan ju ge ett antal uttömmande stickprovsvariabler, t ex utgör ju det ursprungliga stickprovet en sådan. Om observationerna är oberoende likafördelade utgör också ordningsvariablerna uttömmande stickprovsvariabler. Detta innebär att man skulle vilja konstruera en minimal uttömmande stickprovsvariabel, dvs en som är sådan att motsvarande partition av $\Omega$ inte kan "klumpas ihop" ytterligare och ändå vara uttömmande. Detta avsnitt orienterar lite om hur detta kan gå till.

Låt

$$D(\boldsymbol{x}) = \{\boldsymbol{y} \in R^n : f(\boldsymbol{y};\theta) = k(\boldsymbol{y},\boldsymbol{x}) f(\boldsymbol{x};\theta)\}$$

där alltså $k(\boldsymbol{x},\boldsymbol{y})$ ej beror av $\theta$. Man kan se $D(\boldsymbol{x})$ som mängden av de $\boldsymbol{y}$ för vilka kvoten mellan tätheterna i $\boldsymbol{x}$ och $\boldsymbol{y}$ ej beror av $\theta$. Man övertygar sig rätt lätt om att denna konstruktion är en ekvivalensrelation, dvs en relation $\sim$ mellan punkterna (i detta fall i $R^n$) som uppfyller villkoren

1) $\boldsymbol{x} \sim \boldsymbol{x}$ (reflexivitet)

2) Om $\boldsymbol{x} \sim \boldsymbol{y}$ så gäller $\boldsymbol{y} \sim \boldsymbol{x}$ (symmetri)

3) Om $\boldsymbol{x} \sim \boldsymbol{y}$ och $\boldsymbol{y} \sim \boldsymbol{z}$ så gäller $\boldsymbol{x} \sim \boldsymbol{z}$ (transitivitet)

En ekvivalensrelation delar upp mängden i disjunkta ekvivalensklasser. Med urvalsaxiomet kan vi välja en representant $\boldsymbol{x}_D$ från varje ekvivalensklass. Vi låter nu $G(\boldsymbol{x})$ vara just denna representant $\boldsymbol{x}_D$ från den ekvivalensklass som $\boldsymbol{x}$ tillhör. $G$ definierar en partition av $R^n$. Om vi nu betraktar $G(\boldsymbol{X})$ så är detta en stickprovsvariabel som alltså definierar en partition av $\Omega$. Vi har nu för likelihooden

$$f(\boldsymbol{x},\theta) = k(\boldsymbol{x},\boldsymbol{x}_D) f(\boldsymbol{x}_D,\theta) = k(\boldsymbol{x},G(\boldsymbol{x})) f(G(\boldsymbol{x}),\theta).$$

Men detta är just en sådan faktorisering som innebär att $G(\boldsymbol{X})$ är en uttömmande stickprovsvariabel för $\theta$. Ekvivalensrelationen ger alltså upphov till en uttömmande stickprovsvariabel. Vi vill nu visa att den är minimal, dvs ej kan reduceras ytterligare. Antag därför att vi har en annan uttömmande stickprovsvariabel $t(\boldsymbol{x})$ med partitioner $\Pi$. Vi vill alltså visa att varje medlem i partitionen definierad av $t(\boldsymbol{x})$ helt ligger i någon medlem $D$ av partitionerna definierade av $G(\boldsymbol{x})$. Låt $\boldsymbol{x}$ och $\boldsymbol{y}$ vara punkter som ligger i samma partition definierad av $t$ så att $t(\boldsymbol{x}) = t(\boldsymbol{y})$. Eftersom $t(\boldsymbol{X})$ är uttömmande för $\theta$ gäller att

$$f(\boldsymbol{x},\theta) = r(\boldsymbol{x}) s(t(\boldsymbol{x}),\theta)$$

och (eftersom $t(\boldsymbol{x}) = t(\boldsymbol{y})$)

$$f(\boldsymbol{y},\theta) = r(\boldsymbol{y}) s(t(\boldsymbol{y}),\theta) = r(\boldsymbol{y}) s(t(\boldsymbol{x}),\theta).$$

Detta ger om vi löser ut $s(t(\boldsymbol{x}),\theta)$ ur första relationen

$$f(\boldsymbol{y},\theta) = r(\boldsymbol{y}) \frac{f(\boldsymbol{x},\theta)}{r(\boldsymbol{x})} = k(\boldsymbol{x},\boldsymbol{y}) f(\boldsymbol{x},\theta),$$

med $k(\boldsymbol{x},\boldsymbol{y}) = r(\boldsymbol{y})/r(\boldsymbol{x})$. Alltså gäller att $\boldsymbol{x} \sim \boldsymbol{y}$, eller annorlunda uttryckt att de ligger i samma ekvivalensklass i partitionen definierad av $G$. Alltså är $G(\boldsymbol{X})$ en minimal uttömmande stickprovsvariabel som inte kan reduceras ytterligare. $\square$

Exempel: Om vi har binomialfördelning, dvs $n$ oberoende likafördelade data med sannolikhetsfunktion $\theta^x(1-\theta)^x$ för $x = 0, 1$ så erhåller vi

$$P(\boldsymbol{X} = \boldsymbol{x}; \theta) = \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}.$$

Vi får då

$$\frac{P(\boldsymbol{X} = \boldsymbol{x}; \theta)}{P(\boldsymbol{X} = \boldsymbol{y}; \theta)} = (\frac{\theta}{1-\theta})^{\sum_1^n x_i - \sum_1^n y_i},$$

som inte beror av $\theta$ precis då $\sum_1^n x_i = \sum_1^n y_i$ vilket betyder att $\sum_1^n X_i$ är minimal uttömmande stickprovsvariabel. □

Exempel: Om vi har $n$ oberoende normalfördelade variabler $X_i$ som är $\mathrm{N}(\mu, \sigma)$. Kvoten mellan tätheterna i $\boldsymbol{x}$ och $\boldsymbol{y}$ blir då

$$\frac{f(\boldsymbol{x}; \mu, \sigma)}{f(\boldsymbol{y}; \mu, \sigma)} =$$
$$\exp(-\frac{1}{2\sigma^2}(\sum_1^n (x_i - \mu)^2 - \sum_1^n (y_i - \mu)^2)) =$$
$$\exp\Bigg( - \frac{1}{2\sigma^2}[\sum_1^n (x_i^2 - y_i^2) - 2\mu(\sum_1^n x_i - \sum_1^n y_i)]\Bigg).$$

Denna kvot beror av parametern $(\mu, \sigma)$ om inte både $\sum_1^n x_i = \sum_1^n y_i$ och $\sum_1^n x_i^2 = \sum_1^n y_i^2$. Alltså utgör paret $(\sum_1^n X_i, \sum_1^n X_i^2)$ den minimala uttömmande stickprovsvariabeln. Dessa bestämmer entydigt $\bar{X}$ och stickprovsvariansen $S^2$ så även paret $(\bar{X}, S^2)$ är minimal uttömmande. □

En skattning med minimal varians måste vara en funktion av minimala uttömmande stickprovsvariabler. Vi skulle annars med Rao-Blackwells sats kunna konstruera en ny skattning med mindre varians som är en funktion av den minimala uttömmande.