# Föreläsning 10

Statistisk inlärning och dataanalys (Kungliga Tekniska Högskolan)

# Föreläsning 10

I föregående föreläsningen utforskade vi en ganska generell familj av hypotestester som har formen

$$\text{Förkasta } H_0 \text{ om } \lambda(\boldsymbol{x}) \leq c$$

där $\lambda(\boldsymbol{x})$ är likelihoodkvoten och $c \in (0,1)$. I sådana tester behöver vi att bestämma oss vilken $c \in (0,1)$ att använda. Vi ska presentera ett naturligtvis sätt för att bestämma $c$ men först ska vi betrakta ett visst typ av test som kan dyka upp inom familjen av likelihood-kvottester.

## 10.1 $t$-tester

Eftersom likelihood-kvottester kräver att vi maximera $L(\theta|\boldsymbol{x})$ över hela $\Omega_0$ kan vi använda dem för att konstruera tester som involverar andra parametrar (som kallas för *störparametrar*). Störparameter påverkar inte metoden för att konstruera testet men de kan påverka resultatet.

**Exempel 10.1.** Låt $x_1, \ldots, x_n$ vara ett stickprov från en $\mathrm{N}(\mu, \sigma^2)$-population där både $\mu \in \mathbb{R}$ och $\sigma^2 \geq 0$ är okända. Vi skulle vilja testa

$$H_0 : \mu \leq \mu_0 \qquad \text{mot} \qquad H_1 : \mu > \mu_0.$$

Notisera att vårt test gäller bara $\mu$. Eftersom testet sätter inga begränsningar på $\sigma^2$ är $\sigma^2$ en störparameter och testet kan skrivas som

$$H_0 : \mu \leq \mu_0 \text{ och } \sigma^2 \geq 0 \qquad \text{mot} \qquad H_1 : \mu > \mu_0 \text{ och } \sigma^2 \geq 0.$$

Sedan har vi parameterrummet

$$\Omega_0 = \{(\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 \geq 0\} = (-\infty, \mu_0] \times [0, \infty)$$

under $H_0$. Vi såg i Exempel 5.2 att ML-skattningen över hela parameterrummet $\Omega$ är $\hat{\mu} = \bar{x}$ och $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n-1}{n}s^2$. Det följer att

$$\begin{aligned}
\sup_{\mu \in \mathbb{R}, \sigma^2 \geq 0} L(\mu, \sigma^2 | \bar{x}) &= L(\hat{\mu}, \hat{\sigma}^2 | \boldsymbol{x}), \\
&= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n (x_i - \hat{\mu})^2}, \\
&= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n (x_i - \bar{x})^2}, \\
&= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-\frac{1}{2\hat{\sigma}^2} n\hat{\sigma}^2}, \\
&= \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-\frac{n}{2}}.
\end{aligned}$$

För $\Omega_0$ är ML-skattningen $(\hat{\mu}_0, \hat{\sigma}_0^2)$ detsamma som $(\hat{\mu}, \hat{\sigma}^2)$ när $\bar{x} = \hat{\mu} \leq \mu_0$. I detta fall har vi att

$$\sup_{(\mu, \sigma^2) \in \Omega_0} L(\mu, \sigma^2 | \boldsymbol{x}) = \sup_{\mu \in \mathbb{R}, \sigma^2 \geq 0} L(\mu, \sigma^2 | \boldsymbol{x}) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} e^{-\frac{n}{2}}.$$

Å andra sidan, om $\bar{x} > \mu_0$ har vi att $L(\mu, \sigma^2 | \boldsymbol{x})$ maximeras på $(\mu_0, \hat{\sigma}_0^2)$ där

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2.$$

(Det är bra övning att visa detta!) Det följer att

$$\begin{aligned}\sup_{\mu\le\mu_0,\sigma^2\ge 0} L(\mu,\sigma^2|\boldsymbol{x}) &= L(\mu_0,\hat\sigma_0^2|\boldsymbol{x}),\\ &= \frac{1}{(2\pi\hat\sigma_0^2)^{n/2}} e^{-\frac{1}{2\hat\sigma_0^2}\sum_{i=1}^n (x_i-\mu_0)^2},\\ &= \frac{1}{(2\pi\hat\sigma_0^2)^{n/2}} e^{-\frac{1}{2\hat\sigma_0^2}n\hat\sigma_0^2},\\ &= \frac{1}{(2\pi\hat\sigma_0^2)^{n/2}} e^{-\frac{n}{2}}.\end{aligned}$$

Så har vi likelihood-kvottestvariablen

$$\begin{aligned}\lambda(\boldsymbol{x}) &= \begin{cases} 1 & \bar x \le \mu_0,\\ \frac{L(\mu_0,\hat\sigma_0^2|\boldsymbol{x})}{L(\bar x,\hat\sigma^2|\boldsymbol{x})} & \bar x > \mu_0,\end{cases}\\ &= \begin{cases} 1 & \bar x \le \mu_0,\\ \left(\frac{\hat\sigma^2}{\hat\sigma_0^2}\right)^{n/2} & \bar x > \mu_0.\end{cases}\end{aligned}$$

Eftersom $c < 1$ har vi likelihood-kvottestet

$$\text{Förkasta } H_0: \mu \le \mu_0 \text{ om } \left(\frac{\hat\sigma^2}{\hat\sigma_0^2}\right)^{n/2} \le c \text{ och } \bar x > \mu_0.$$

Med hjälp av lite algebra kan vi formulera om testet i ett bättre sätt. Notisera första att

$$\begin{aligned}\hat\sigma^2 &= \frac{n-1}{n}s^2, \qquad \text{och}\\ \hat\sigma_0^2 &= \frac{1}{n}\sum_{i=1}^n (x_i-\mu_0)^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\bar x)^2 + (\bar x-\mu_0)^2 = \frac{n-1}{n}s^2 + (\bar x-\mu_0)^2.\end{aligned}$$

Det följer att $\left(\frac{\hat\sigma^2}{\hat\sigma_0^2}\right)^{n/2} \le c$ om och endast om

$$\begin{aligned}\frac{\frac{n-1}{n}s^2}{\frac{n-1}{n}s^2 + (\bar x-\mu_0)^2} &\le c^{2/n}, \qquad \text{om och endast om}\\ \frac{n-1}{n-1+\frac{(\bar x-\mu_0)^2}{s^2/n}} &\le c^{2/n}, \qquad \text{om och endast om}\\ \frac{(\bar x-\mu_0)^2}{s^2/n} &\ge (n-1)\left(\frac{1}{c^{2/n}}-1\right), \qquad \text{om och endast om}\\ \left|\frac{(\bar x-\mu_0)}{s/\sqrt{n}}\right| &\ge c'\end{aligned}$$

där

$$c' = \sqrt{(n-1)\left(\frac{1}{c^{2/n}}-1\right)}.$$

Så har vi testet

$$\text{Förkasta } H_0:\ \mu \le \mu_0 \text{ om } \left|\frac{(\bar x-\mu_0)}{s/\sqrt{n}}\right| \ge c' \text{ och } \bar x > \mu_0.$$

Men detta är lika

$$\text{Förkasta } H_0:\ \mu \le \mu_0 \text{ om } \frac{(\bar x-\mu_0)}{s/\sqrt{n}} \ge c'.$$

Det följer att testvariablen $\lambda(\boldsymbol{X})$ är $t$-fördelad under $H_0$. Därför testet kallas för ett *t-test*.

Generellt sett är ett *t-test* ett test som beror på en testvariabel som är $t$-fördelad under nollhypotesen. Det finns många olika exempel av sådana tester som kan vara användbara. Exempelvis, finns det ett t-test för betingat oberoende för normalfördelad stokastiska variabler:

**Exempel 10.2.** Låt $\boldsymbol{X} = (X_1, \ldots, X_m) \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ och ta $i, j \in [m]$ och $C \subseteq [m] \setminus C$. Vi skulle vilja testa

$$H_0 : X_i \perp\!\!\!\perp X_j \mid \boldsymbol{X}_C \qquad \text{mot} \qquad H_1 : X_i \not\perp\!\!\!\perp X_j | \boldsymbol{X}_C.$$

För att göra det vi ska konstruera ett $t$-test baserad på *partiell korrelation* testvariablen. Inversen av matrisen $\Sigma$, betecknas $K = [k_{ij}]$, kallas *precisionsmatrisen*. *Partiell korrelationen* av $X_i$ och $X_j$ givet $\boldsymbol{X}_C$ är

$$\rho_{i,j|C} = \frac{k_{ij}}{\sqrt{k_{ii}k_{jj}}}.$$

Vi beror inte på variabler i $\boldsymbol{X}_{[m]\setminus(\{i,j\}\cup C)}$ (de ger oss bara störparameter). Därför kan vi marginalisera dem ut och arbete med $\boldsymbol{X}_{\{i,j\}\cup C} \sim \mathrm{N}(\boldsymbol{\mu}_{\{i,j\}\cup C}, \Sigma_{\{i,j\}\cup C,\{i,j\}\cup C})$. I följande låter vi $\boldsymbol{\mu} = \boldsymbol{\mu}_{\{i,j\}\cup C}$ och $\Sigma = \Sigma_{\{i,j\}\cup C,\{i,j\}\cup C}$ för att förenkla notationen.

Eftersom $\Sigma$ är postivt definit följer det att $K$ är också positivt definit och därför $k_{ii} \neq 0$ för alla $i \in [m]$. Vi har också att $\rho_{i,j|C} = 0$ om och endast om $k_{ij} = 0$. Vi kan använda *Cramers Regel* från linjär algebra för att beräkna $k_{ij}$ som

$$k_{ij} = \frac{(-1)^{i+j}\det(\Sigma_{C\cup\{i\},C\cup\{j\}})}{\det(\Sigma)}.$$

Eftersom $\Sigma$ är postivt definit har vi att $\det(\Sigma) \neq 0$. Vi har också att $k_{ij} = 0$ om och endast om $\det(\Sigma_{C\cup\{i\},C\cup\{j\}}) = 0$. Vi såg i Sats 3.1 att $X_i \perp\!\!\!\perp X_j|\boldsymbol{X}_C$ om och endast om $\Sigma_{C\cup\{i\},C\cup\{j\}}$ har rang $|C|$. Eftersom $\Sigma_{C\cup\{i\},C\cup\{j\}}$ har submatrisen $\Sigma_{C,C}$ som är rang $|C|$ (eftersom det är postivt definit) följer det att $\Sigma_{C\cup\{i\},C\cup\{j\}}$ har rang $|C|$ om och endast om $\det(\Sigma_{C\cup\{i\},C\cup\{j\}}) = 0$. Vi har bevisat att $\rho_{i,j|C} = 0$ om och endast om $X_i \perp\!\!\!\perp X_j|\boldsymbol{X}_C$.

Givet ett stickprov $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ från $\mathrm{N}(\boldsymbol{\mu}, \Sigma)$ definieras vi *stickprovspartiell korrelationen*

$$r_{i,j|C} = \frac{\hat{k}_{ij}}{\sqrt{\hat{k}_{ii}\hat{k}_{jj}}}$$

där $\hat{K} = [\hat{k}_{ij}] = \hat{\Sigma}^{-1}$ är *stickprovsprecisionsmatrisen*; dvs. inversen av stickprovscovariansmatrisen från Sats 5.1. Om ML-skattningen $\hat{\Sigma}$ för $\Sigma$ existeras har vi att $r_{i,j|C}$ är en ML-skattningen för $\rho_{i,j|C}$ med hjälp av Sats 5.2. Så ett bra test för $H_0 : X_i \perp\!\!\!\perp X_j|\boldsymbol{X}_C$ är att förkasta $H_0$ om $r_{i,j|C}$ är lång borta från 0. Det kan visas att testvariablen

$$T(\boldsymbol{X}) = \sqrt{n - |C| - 2}\,\frac{r_{i,j|C}}{\sqrt{1 - r_{i,j|C}^2}}$$

är $t(N - |C| - 2)$-fördelad. Det kan också visas (kanske i en kurs på mastersnivå om regression) att $-1 \leq r_{i,j|C} \leq 1$. Därför $T(\boldsymbol{X})$ ska vara liten eller stor när $r_{i,j|C}$ är liten respektive stor.

Så vi kan använda $T(\boldsymbol{X})$ för att formulera ett $t$-test som

$$\text{Förkasta } H_0 : X_i \perp\!\!\!\perp X_j|\boldsymbol{X}_C \text{ om } |T(\boldsymbol{x})| \geq c.$$

**Exempel 10.3.** Vi kan tillämpa $t$-testet från Exempel 10.2 för att lära oss en hierarki om normalfördelade variabler.

Till exempel anta att vi har $\boldsymbol{X} = (X_1, X_2, X_3)$ där $X_i$ är uttrycksnivån för gene $i$. Vi har 100 utfall från $\boldsymbol{X}$:

$$
\begin{array}{cccc}
(1.79, -1.26, 1.69) & (1.49, 0.73, 1.4) & (-1.34, 1.01, -0.48) & (-0.67, -0.78, -0.29) \\
(-1.35, -2.99, 0.61) & (-1.56, 0.08, -1.12) & (0.97, -0.78, 1.09) & (1.55, -0.66, 1.88) \\
(0.38, 0.68, 0.21) & (0.98, -0.61, 0.81) & (0.53, -0.12, 1.15) & (-0.81, 1.37, -0.64) \\
(0.13, 0.92, 0.04) & (-0.09, -1.73, 0.59) & (1.73, -0.53, 0.19) & (2.43, -1.96, 2.54) \\
(-0.4, -0.97, -0.64) & (-0.71, 3.33, -3.04) & (0.16, 1.04, -0.99) & (-1.63, 1.53, -1.16) \\
(1.0, -1.6, 0.81) & (-0.49, -3.43, 1.53) & (1.21, 0.5, -1.55) & (0.76, -0.9, 0.04) \\
(0.8, -1.93, 1.3) & (0.25, -2.0, 2.04) & (0.33, -0.07, 0.4) & (0.23, 0.65, -0.29) \\
(-0.98, 0.1, -0.71) & (-0.08, -1.11, 0.57) & (-0.22, -1.28, 0.56) & (0.48, 1.79, -0.68) \\
(0.33, -0.27, 0.13) & (-0.33, -2.31, 2.14) & (0.89, -1.24, 1.12) & (-0.61, 0.41, 0.87) \\
(0.89, 0.99, -0.63) & (-0.15, -0.89, 0.62) & (-1.0, -0.79, -0.2) & (1.41, -0.13, 0.6) \\
(1.1, -2.21, 2.81) & (-1.36, 0.67, -2.06) & (0.32, 1.24, -0.27) & (-0.34, 2.52, -1.17) \\
(0.1, -1.29, 0.74) & (0.44, -1.73, 1.84) & (-0.75, -0.26, 1.35) & (-0.62, 1.8, -1.53) \\
(1.48, -2.26, 1.93) & (0.37, -0.55, 1.29) & (-0.75, 1.75, -0.27) & (1.55, -0.15, 0.66) \\
(0.47, -2.65, 0.91) & (1.45, -2.17, 1.56) & (1.15, -0.69, -0.64) & (-1.31, -0.5, 0.23) \\
(-1.31, -0.38, 0.39) & (-1.54, 0.21, -1.74) & (-0.22, 0.2, 0.86) & (0.0, -1.86, 0.75) \\
(-0.39, 1.66, -0.33) & (-0.09, -2.56, 2.04) & (-1.23, 3.13, -3.17) & (0.72, 0.08, 0.5) \\
(0.59, 0.4, -0.24) & (1.19, -0.83, 1.08) & (-1.67, 1.27, -1.76) & (-0.77, 1.75, -2.12) \\
(1.1, -1.61, 1.55) & (0.62, -1.58, 1.85) & (-0.43, 0.65, 0.37) & (-0.15, -1.08, 1.44) \\
(0.16, -0.72, 1.02) & (0.49, 0.38, -0.41) & (-1.04, -3.17, 0.5) & (-0.87, -0.24, -0.33) \\
(-0.11, 2.29, -1.0) & (-1.43, 0.62, -0.59) & (0.47, 0.64, -0.42) & (-1.25, 0.98, -0.48) \\
(-1.27, 0.06, -0.88) & (0.62, 0.13, 1.35) & (0.51, -1.53, 0.09) & (-1.0, -1.03, 0.71) \\
(-0.51, -0.18, -0.68) & (-0.32, -0.28, 0.02) & (0.94, 1.18, -1.08) & (-1.02, -0.71, -1.72) \\
(1.81, -0.84, 0.48) & (-0.43, 2.0, -0.35) & (1.28, -0.84, 0.94) & (-1.02, -0.68, 0.68) \\
(1.66, -3.36, 3.01) & (-0.08, -1.42, 1.35) & (-0.49, 0.35, 0.49) & (-0.64, 0.77, -0.81) \\
(0.77, 1.73, -0.93) & (0.4, 0.98, 0.55) & (1.58, -0.53, 1.34) & (-1.52, 1.77, -0.67)
\end{array}
$$

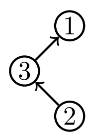
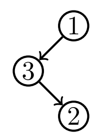
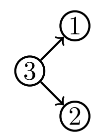
Här indikerar en negativ uttrycksnivå hämning i förhållande till en normaliserad baslinje. Covariansmatrisen för datan är

$$
\hat{\Sigma} = \begin{bmatrix} 0.95744706 & -0.43315334 & 0.63362794 \\ -0.43315334 & 2.02951292 & -1.29245954 \\ 0.63362794 & -1.29245954 & 1.44633216 \end{bmatrix}.
$$

Vi testar alla möjliga betingat oberoenderelationer med $t$-testet från Exempel 10.2 med $c = 1$.

| $H_0$ | $T(\boldsymbol{x})$ |
|---|---|
| $X_1 \perp\!\!\!\perp X_2$ | $-1.464$ |
| $X_1 \perp\!\!\!\perp X_3$ | $3.223$ |
| $X_2 \perp\!\!\!\perp X_3$ | $-7.440$ |
| $X_1 \perp\!\!\!\perp X_2 \vert X_3$ | $0.732$ |
| $X_1 \perp\!\!\!\perp X_3 \vert X_2$ | $2.635$ |
| $X_2 \perp\!\!\!\perp X_3 \vert X_1$ | $-6.535$ |

Testet förkastar alla hypoteser utanför $X_1 \perp\!\!\!\perp X_2|X_3$. Så vi drar slutsatsen från Sats 3.6 att fördelningen är Markovsk till följande grafer:

ML-skattningen för $(\boldsymbol{\mu}, \Sigma)$ är

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} 0.0571 \\ -0.2389 \\ 0.2354 \end{bmatrix} \qquad \text{och} \qquad \hat{\Sigma} = \begin{bmatrix} 0.95744706 & -0.43315334 & 0.63362794 \\ -0.43315334 & 2.02951292 & -1.29245954 \\ 0.63362794 & -1.29245954 & 1.44633216 \end{bmatrix}.$$

Om vi skatta fördelningen som $\mathrm{N}(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ säger den sista grafen att vi kan modellera fördelningen med hierarkin:

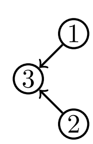
$$\begin{aligned} X_1 &\sim \mathrm{N}(0.057, 0.957), \\ X_2|X_1 = x_1 &\sim \mathrm{N}(-0.691x_1 - 0.213, 1.834), \\ X_3|X_1 = x_1 &\sim \mathrm{N}(0.897x_1 + 0.198, 1.027). \end{aligned}$$

Här beräknade vi väntevärdena och varianserna med hjälp av Sats 2.3.

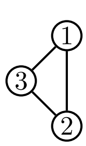
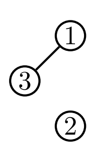
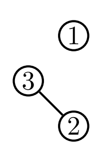
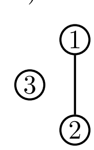
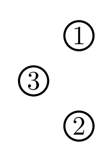
I Exempel 10.3 testade vi hypoteserna med $c = 1$. Eftersom $T(\boldsymbol{X}) \sim t(100 - |C| - 2)$ för varje test har vi att $P(|T(\boldsymbol{X})| \geq 1) \approx 0.3$ vilken är kanske lite stor. Om vi gjorde testarna med $c = 2$ istället har vi att $P(|T(\boldsymbol{X})| \geq 2) \approx 0.05$ (detsamma sannolikheten som vi använde i Exempel 9.1). Under ett sådant test förkastar vi $H_0$ med väldigt liten risk av fel men vi accepterar då båda

$$H_0 : X_1 \perp\!\!\!\perp X_2 \qquad \text{och} \qquad H_0 : X_1 \perp\!\!\!\perp X_2|X_3.$$

Det finns egentligen inte en enda graf som kodar båda $X_1 \perp\!\!\!\perp X_2$ och $X_1 \perp\!\!\!\perp X_2|X_3$ i deras d-separationer (bevisa det!). Så $X_1 \perp\!\!\!\perp X_2|X_3$ föreslår att bra hierarkier är graferna i Exempel 10.3 medan $X_1 \perp\!\!\!\perp X_2$ föreslår att vi borde använda hierarkin



Då har vi fyra olika modeller som kommer ifrån två olika Markovekvivalensklasser. Med våra tester har vi uteslutit alla graferna i fem Markovekvivalensklasser – de med skelett (och v-strukturer):

Vi kan dock inte bestämma vilken Markovekvivalensklass av de två givna av $X_1 \perp\!\!\!\perp X_2$ respektive $X_1 \perp\!\!\!\perp X_2|X_3$ är det bästa valet baserad på bara våra hypotesttester. Detta är ett exempel av problemet av *modellval*. I de kommande föreläsningen kommer vi se olika metoder för modellval som kan hjälpa oss bestämma mellan dessa modeller.

## 10.2 Styrkfunktionen

Hittills har vi hittat likelihood-kvottester i formen

$$\text{Förkasta } H_0 \text{ om } \lambda(\boldsymbol{x}) \leq c$$

för någon $c \in (0,1)$. I våra exemplar antingen lämnade vi värdet av $c$ ospecificerat eller valde c utan förklarning (varför tog vi $c = 0.27$ för vulkanutbrottningsdatan i Exempel 9.3?). Nu betraktar vi en funktion vilken ska ge oss ett sätt för att utvärdera testerna för olika $c$ och välja $c$ för att uppnå önskad prestationsnivå.

*Styrkfunktionen* av ett test med kritiskt området $R$ ges av

$$\beta(\theta) = P(\boldsymbol{X} \in R|\theta).$$

Styrkfunktionen därför mäter för oss sannolikheten av förkastning av nollhypotesen för olika värden på $\theta \in \Omega$ där $\Omega$ är hela parameterrummet. Vi skulle företrädesvis vilja ha att vi förkasta $H_0$ när $\theta \in \Omega_1$ och acceptera $H_0$ när $\theta \in \Omega_0$. Vi vill därför ha att $\beta(\theta)$ är låg när $\theta \in \Omega_0$ och $\beta(\theta)$ är hög för $\theta \in \Omega_1$.

Om vårt test är en punkttest - exempelvis

$$\text{Förkasta } H_0 : \theta = \theta_0 \text{ om } \lambda(\boldsymbol{x}) \leq c$$

är kritiskt området $R = \{\theta_0\}$ och därför är $\beta(\theta_0) = P(\lambda(\boldsymbol{X}) \leq c|\theta_0)$. Vi såg i Föreläsning 9 ett sådant exempel med vårt test för leveranshästigheten av en viss fabrik där vi hade $\Omega_0 = \{3\}$ och kritiskt området $R = \{\boldsymbol{x} : \boldsymbol{x} \geq t\}$. För detta test specificerade vi $\beta(3) = P(\boldsymbol{X} \geq t|3) = 0.05$; det vill säga att vi bestämde oss en *signifikansnivå* av 0.05. Givet att vår generell familj av tester – likelihood-kvottesterna – specificerar kritiskt området igenom värdet av $c \in (0,1)$ skulle vi vilja ha ett likadant sätt att bestämma oss vilken $c$ att använda för att ha en specificerad signifikansnivå. Det kan vi göra med hjälp av styrkfunktionen men vi kan göra även bättre eftersom $\beta(\theta)$ också kodar informationen om sannolikheten att vi acceptera nollhypotesen trots att det är falskt.

När vi felaktigt förkastar nollhypotesen (dvs. vi förkastar $H_0$ trots att det stämmer) kallas det för ett *typ-I fel*. Å andra sidan, när vi felaktigt förkastar alternativhypotesen (dvs. felaktigt acceptera $H_0$) kallas det för ett *typ-II fel*. Dessa är de enda två typer av fel som vi kan göra i vårt test:

| | | Beslut: acceptera $H_0$ | Beslut: acceptera $H_1$ |
|---|---|---|---|
| Sanning | $H_0$ | korrekt | Typ-I fel |
| Sanning | $H_1$ | Type-II fel | korrekt |

Låt $R^c$ beteckna komplementet av kritiskt området $R$. Vårt test förkasta $H_0$ om och endast om $\boldsymbol{x} \in R$. Så det följer att $P(\boldsymbol{X} \in R|\theta)$ är sannolikheten av ett Typ-I fel när $\theta \in \Omega_0$. När $\theta \in \Omega_1$ har vi att $P(\boldsymbol{X} \in R^c|\theta)$ är sannolikheten av ett Typ-II fel där. Eftersom $P(\boldsymbol{X} \in R^c|\theta) = 1 - P(\boldsymbol{X} \in R|\theta)$ för en viss $\theta \in \Omega$ har vi enklare att

$$\beta(\theta) = P(\boldsymbol{X} \in R|\theta) = \begin{cases} \text{sannolikheten av ett Typ-I fel} & \theta \in \Omega_0, \\ \text{ett minus sannolikheten av ett Typ-II fel} & \theta \in \Omega_1. \end{cases}$$

Detta förklara varför vi skulle vilja ha $\beta(\theta)$ låg när $\theta \in \Omega_0$ och hög när $\theta \in \Omega_1$!

**Exempel 10.4.** Låt $x_1, \ldots, x_n$ vara ett stickprov från en $\mathrm{N}(\theta, 1)$-population med 1 känt. I Exempel 9.2 såg vi att likelihood-kvottestet för

$$H_0 : \theta = \theta_0 \qquad \text{mot} \qquad H_1 : \theta \neq \theta_0$$

ges av

$$\text{Förkasta } H_0 \text{ om } |\bar{x} - \theta_0| \geq c$$

där $c > 0$. Kritiskt området för ett visst $c$ är då $R = \{\boldsymbol{x} : |\bar{x} - \theta_0| \geq c\}$. Så får vi styrkfunktionen

$$\begin{aligned} \beta(\theta) &= P(|\bar{x} - \theta_0| \geq c|\theta), \\ &= P(\bar{x} - \theta_0 \geq c|\theta) + P(\bar{x} - \theta_0 \leq -c|\theta). \end{aligned}$$

Vi vet från Sats 4.1 att $\bar{X} \sim \mathrm{N}(\theta, 1/n)$ och därför har vi att

$$\begin{aligned}
\beta(\theta) &= P\left(\bar{x} \geq c + \theta_0 | \theta\right) + P\left(\bar{x} \leq \theta_0 - c | \theta\right), \\
&= P\left(\frac{\bar{X} - \theta}{1/\sqrt{n}} \geq \frac{(\theta_0 - \theta) + c}{1/\sqrt{n}} | \theta\right) + P\left(\frac{\bar{X} - \theta}{1/\sqrt{n}} \leq \frac{(\theta_0 - \theta) - c}{1/\sqrt{n}} | \theta\right), \\
&= P\left(Z \geq \frac{(\theta_0 - \theta) + c}{1/\sqrt{n}}\right) + P\left(Z \leq \frac{(\theta_0 - \theta) - c}{1/\sqrt{n}}\right), \\
&= 1 - F_Z(\sqrt{n}((\theta_0 - \theta) + c)) + F_Z(\sqrt{n}((\theta_0 - \theta) - c)),
\end{aligned}$$

där $Z \sim \mathrm{N}(0, 1)$ och $Z$ har fördelningsfunktion $F_Z(z)$. Anta nu att vi skulle vilja ha att sannolikheten av ett Typ-I fel är $\alpha \in (0, 1)$; dvs.

$$\beta(\theta_0) = \alpha.$$

Eftersom $f_Z(z)$ är symmetrisk runt 0 har vi då att

$$\begin{aligned}
\alpha &= \beta(\theta_0), \\
\alpha &= 1 - F_Z(\sqrt{n}((\theta_0 - \theta_0) + c)) + F_Z(\sqrt{n}((\theta_0 - \theta_0) - c)), \\
\alpha &= 1 - F_Z(\sqrt{n}(c)) + F_Z(-\sqrt{n}(c)), \\
\alpha &= 2(1 - F_Z(c\sqrt{n})), \\
F_Z(c\sqrt{n}) &= 1 - \frac{\alpha}{2},
\end{aligned}$$

vilken händer om och endast om $\lambda_{\alpha/2} = c\sqrt{n}$ där $\lambda_t$ betecknas $t$-kvantilen av $Z \sim \mathrm{N}(0, 1)$; dvs. $\lambda_t$ är värdet på $z$ som ger att sannolikheten $F_Z(z) = P(Z \leq \lambda_t) = 1 - t$. Så följer det att

$$c = \frac{\lambda_{\alpha/2}}{\sqrt{n}}.$$

Till exempel, om vi har 100 utfall i stickprovet och vi skulle vilja ha en sannolikhet av ett Typ-I fel på $\alpha = 0.05$ tar vi

$$c = \frac{\lambda_{0.025}}{10} = \frac{1.96}{10} = 0.196.$$

På liknande sätt kan vi hitta värdet för $c$ i en likelihood-kvottest som ger oss en viss maximal sannolikhet för ett Typ-I fel när $\Omega_0$ innehåller mer än ett element.

**Exempel 10.5.** Låt $x_1, \ldots, x_n$ vara ett stickprov från en $\mathrm{N}(\theta, \sigma^2)$-population där $\sigma^2$ är känt. Man kan visa att likelihood-kvottestet för

$$H_0 : \theta \leq \theta_0 \qquad \text{mot} \qquad H_1 : \theta > \theta_0$$

ges av

$$\text{Förkasta } H_0 \text{ om } \bar{x} - \theta_0 > c$$

för någon $c > 0$. På samma sätt som det föregående exemplet har vi styrkfunktionen

$$\begin{aligned}
\beta(\theta) &= P(\bar{x} - \theta_0 > c | \theta), \\
&= P\left(\frac{\bar{X} - \theta}{\sigma/\sqrt{n}} > \frac{c + \theta_0 - \theta}{\sigma/\sqrt{n}} | \theta\right), \\
&= P\left(Z > \frac{c + \theta_0 - \theta}{\sigma/\sqrt{n}} | \theta\right), \\
&= 1 - F_Z\left(\frac{c + \theta_0 - \theta}{\sigma/\sqrt{n}}\right),
\end{aligned}$$

där $Z \sim \mathrm{N}(0, 1)$ och $Z$ har fördelningsfunktion $F_Z(z)$.

I detta exempel är $\Omega_0 = \{\theta : \theta \leq \theta_0\}$ (inte bara $\{\theta\}$). Så för att minimera sannolikheten av ett Typ-I fel skulle vi ha att det maximala värdet av $\beta(\theta)$ över $\Omega_0$ är liten. Om vi skulle vilja ha denna sannolikhet lika med $\alpha$ vill vi då ha att

$$\begin{aligned}
\alpha &= \max_{\theta \leq \theta_0} \beta(\theta), \\
&= \max_{\theta \leq \theta_0} \left(1 - \Phi\left(\frac{c + \theta_0 - \theta}{\sigma/\sqrt{n}}\right)\right).
\end{aligned}$$

Eftersom $\Phi(z)$ är strikt växande (det är en fördelningfunktion) är maximalt av $\beta(\theta)$ uppnådde på $\theta = \theta_0$. Så har vi att

$$\begin{aligned}\alpha &= \max_{\theta \le \theta_0} \beta(\theta), \\ &= 1 - \Phi\left(\frac{c + \theta_0 - \theta_0}{\sigma/\sqrt{n}}\right), \\ &= 1 - \Phi\left(\frac{c}{\sigma/\sqrt{n}}\right),\end{aligned}$$

vilken händer om och endast om $\lambda_\alpha = \frac{c}{\sigma/\sqrt{n}}$. Därför skulle vi vilja ta

$$c = \frac{\lambda_\alpha \sigma}{\sqrt{n}}.$$

Därmed är önskat test

$$\text{Förkasta } H_0 : \theta \le \theta_0 \text{ om } \bar{x} > \theta_0 + \tfrac{\lambda_\alpha \sigma}{\sqrt{n}}.$$

### 10.2.1 Storlek och nivå

Att välja ett test genom att införa begränsningar på sannolikheten av ett Typ-I fel som i Exemplar 10.4 och 10.5 är ganska vanligt. Vi säger att ett test har *storlek* $\alpha$ om

$$\max_{\theta \in \Omega_0} \beta(\theta) = \alpha.$$

Vanliga val av storleker för testar inkluderar $\alpha = 0.05, 0.01, 0.001$. Ibland kan vi inte garantera att testets storlek är exakt lika med $\alpha$. I sådana fall kan vi begränsar värdet av styrkfunktionen över $\Omega_0$ med $\alpha$ istället. Vi säger att ett test har *nivå* $\alpha$ om

$$\max_{\theta \in \Omega_0} \beta(\theta) \le \alpha.$$

När ett test är ett punkttest (dvs., $\Omega_0 = \{\theta_0\}$ innehåller bara ett element) har testet nivå $\alpha$ när det har storlek $\alpha$. Det följer eftersom

$$\max_{\theta \in \Omega_0} \beta(\theta) = \beta(\theta_0) = \max_{\theta \in \Omega_0} \beta(\theta).$$

Till exempel om vi låter $c = \lambda_{\alpha/2}/\sqrt{n}$ har testet i Exempel 10.4 nivå och storlek $\alpha$. Eftersom $\Omega = \{\theta_0\}$ kan vi se detta genom att kontrollera $\beta(\theta_0) = P(|\bar{x} - \theta_0| \ge \lambda_{\alpha/2}/\sqrt{n}|\theta_0) = \alpha$. I detta fall behövs ingen maximering – till skillnad från Exempel 10.5. Diskussionen i Exempel 10.5 visar att $c = \lambda_\alpha \sigma/\sqrt{n}$ ger ett test av nivå och storlek $\alpha$ eftersom $\beta(\theta)$ är strikt växande på $\theta \le \theta_0$. Generellt sett är ett test av storlek lika med $\alpha$ ett test av nivå $\alpha$ men det omvända dock inte gäller nödvändigtvis.

**Exempel 10.6.** Vi återgår till testet i Exempel 9.3 där vi hade $X_1, \ldots, X_n$ oberoende och $f_X(x|\theta)$-fördelade där $f_X(x|\theta)$ tillhörde en läge-familj med standardfördelningen $Z \sim \text{Exp}(1)$. Likelihood-kvottestet för

$$H_0 : \theta \le \theta_0 \qquad \text{mot} \qquad H_1 : \theta > \theta_0$$

var

$$\text{Förkasta } H_0 : \theta \le \theta_0 \text{ om } \min_i(x_i) - \theta_0 \ge c$$

för någon $c > 0$. Styrkfunktionen för testet är

$$\beta(\theta) = P(\boldsymbol{X} \in R|\theta) = P(\min_i(X_i) - \theta_0 \ge c|\theta) = P(\min_i(X_i) \ge c + \theta_0|\theta).$$

Det följer att $\beta(\theta)$ är en strikt växande funktion eftersom det blir mer sannolikt att $\min_i(X_i)$ är större än $c + \theta_0$ när vi skiftar $\theta$ åt höger. Så har vi att

$$\begin{aligned}\max_{\theta \in \Omega_0} \beta(\theta) &= \max_{\theta \le \theta_0} P(\min_i(X_i) \ge c + \theta_0|\theta), \\ &= P(\min_i(X_i) \ge c + \theta_0|\theta_0).\end{aligned}$$

Eftersom $Y = \min_i(X_i)$ har täthetsfunktion

$$f_Y(y|\theta) = \begin{cases} ne^{-n(y-\theta)} & y \geq \theta, \\ 0 & y < \theta, \end{cases}$$

får vi att

$$\begin{aligned} P(\min_i(X_i) \geq c + \theta_0|\theta_0) &= \int_{c+\theta_0}^{\infty} ne^{-n(y-\theta_0)}dy, \\ &= e^{-n(c+\theta_0-\theta_0)}, \\ &= e^{-nc}. \end{aligned}$$

Om vi löser

$$e^{-nc} = \max_{\theta \in \Omega_0} \beta(\theta) = \alpha$$

för $c$ får vi $c = -\frac{\log(\alpha)}{n}$. Så har testet

$$\text{Förkasta } H_0 : \theta \leq \theta_0 \text{ om } \min_i(x_i) - \theta_0 \geq -\tfrac{\log(\alpha)}{n}$$

storlek $\alpha$. I vårt exempel med vulkanutbrottnings data tog vi $c = 0.27$. Det följer att testet hade storlek $\alpha = e^{-(11)(0.27)} \approx 0.05$ vilken är en ganska rimligt storlek för ett test.

### 10.2.2 Styrka

När vi har fixat ett storlek för testet har vi begränsat sannolikheten av ett Typ-I fel. Sannolikheten av ett Typ-II fel är dock forfarande obegränsad. Vi såg att sannolikheten av ett Typ-II fel är

$$P(\boldsymbol{x} \in R^c|\theta) = 1 - P(\boldsymbol{x} \in R|\theta) = 1 - \beta(\theta) \qquad (\text{där } \theta \in \Omega_1).$$

Vi säger att ett test har *styrka*

$$\beta(\theta) = P(\boldsymbol{x} \in R|\theta)$$

mot $\theta \in \Omega_1$. Om $\beta(\theta)$ är hög på ”meningsfulla” värden av $\theta$ ska sannolikheten av ett Typ-II fel vara låg för sådana värden. Här måste vi bestämma själv vad är meningsfulla värden av parametern. Till exempel, i Exempel 10.5 testar vi

$$H_0 : \theta \leq \theta_0 \qquad \text{mot} \qquad H_1 : \theta > \theta_0$$

med $\alpha$-nivå testet

$$\text{Förkasta } H_0 : \theta \leq \theta_0 \text{ om } \bar{x} - \theta_0 \geq \lambda_\alpha \sigma/\sqrt{n}.$$

Om vi tar, exempelvis, $\theta_0 = 0$ är det möjligt att det riktiga värdet av $\theta$ är $0.5 \in \Omega_1$. Om vi förkasta $H_0$ eftersom $\bar{x} = 0.025$ kommer vi inte vara överaskad. Å andra sidan, om vi förkasta $H_0$ eftersom $\bar{x} = 7$ är det kanske mer intressant eftersom stickprovsmedelvärdet är ganska lång bort från nollhypotesen. Det kan nog alla hålla med om att $\theta = 7 \in \Omega_1$ är meningsfull medan vi kanske inte alla är överens att det stämmer också för $\theta = 0.025$.

Givet ett nivå $\alpha$ test skulle vi vilja ha att testet har en hög stryka mot meningsfulla värden av $\theta \in \Omega_1$. En vanligtvis nedre gräns för $\beta(\theta)$ är 0.80. När $\beta(\theta) \geq 0.80$ man brukar säger bara att testet har *styrka* mot $\theta$.

**Exempel 10.7.** Låt $X_1, \ldots, X_n$ oberoende och $\mathrm{N}(\theta, 1)$-fördelade. Vi vill testa

$$H_0 : \theta \leq 0 \qquad \text{mot} \qquad H_1 : \theta > 0.$$

Vi såg i Exempel 10.5 att ett 0.05-nivå test är

$$\text{Förkasta } H_0 : \theta \leq 0 \text{ om } \bar{x} \geq 0.5199/\sqrt{n}.$$

Om vi tar $\theta \in \Omega_1$ har testet stryka mot $\theta$

$$\begin{aligned} \beta(\theta) &= P(\bar{x} \geq 0.5199/\sqrt{n}|\theta), \\ &= P\left(\frac{\bar{x} - \theta}{\sqrt{n}} \geq 0.5199 - \theta\sqrt{n}\right), \\ &= \int_{0.5199-\theta\sqrt{n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \end{aligned}$$

Eftersom $\theta \in \Omega_1$ är $\theta$ positiv. Det följer att styrkan av testet ökar när vi ökar stickprovets storlek, $n$. Vi ser också att $P(\bar{x} \geq 0.5199/\sqrt{n}|\theta) \leq P(\bar{x} \geq 0.5199/\sqrt{n}|\theta')$ för $\theta, \theta' \in \Omega_1$ där $\theta \leq \theta'$. Där kan vi fixa en meningsfull nedre gräns för $\theta \in \Omega_1$ och bestämma hur många utfall vi måste ha i stickprovet för att ha stryka mot alla $\theta' \geq \theta$. Exempelvis, om vi fixar $\theta = 0.5$ har vi att

$$P(\bar{x} \geq 0.5199/\sqrt{n}|\theta) \approx 0.7881$$

för $n = 7$ men

$$P(\bar{x} \geq 0.5199/\sqrt{n}|\theta) \approx 0.8133$$

för $n = 8$. Det följer att testet har stryka mot $\theta \geq 0.5$ när vi har minst 8 utfall i stickprovet.

### 10.2.3 Starkaste tester

**(Detta avsnitt är bara fördjupning och ska inte dyka upp på tentamen!)** Ibland kan det vara svårt att specificera meningsfulla värden av $\theta \in \Omega_1$. I sådana fall är det säkrast att arbete med tester som fungerar bäst över alla alternativ. Efter vi har bestämt oss en viss nivå $\alpha$ för våra test kan vi prova att maximera strykan $\beta(\theta)$ över $\Omega_1$. Ett test som ger oss detta maximum ska ha det högsta styrkan som är möjligt över alla $\alpha$-nivå tester. Ett $\alpha$-nivå test kallas för ett *$\alpha$-nivå likformigt starkaste test* (LST) om dess styrkfunktion $\beta(\theta)$ uppfyller

$$\beta(\theta) \geq \beta'(\theta) \qquad \text{för alla } \theta \in \Omega_1$$

där $\beta'(\theta)$ är styrkfunktionen för vilket annat $\alpha$-nivå test som helst. Sådana $\alpha$-nivå tester, när de existerar, är optimala i den meningen att de minimera sannolikheten av ett Typ-II fel för den givna $\alpha$-nivån.

För att hitta LST:er i fallet av punktnollhypoteser mot punktmothypoteser har vi följande

**Sats 10.1** (Neyman-Pearson Sats)**.** *Låt $X$ har täthetsfunktion (eller sannolikhetsfunktion) $f_X(x|\theta)$ där $\theta$ kan ta en av två möjliga värden: $\theta_0$ eller $\theta_1$. Låt $H_1 : \theta = \theta_0$ och $H_1 : \theta = \theta_1$. Så har vi att testet med kriktiskt området*

$$R_{NP} = \left\{ \boldsymbol{x} : \frac{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta_0)}{f_{\boldsymbol{X}}(\boldsymbol{x}|\theta_1)} \leq \eta \right\}$$

*där $\eta$ är valt så att $P(\boldsymbol{x} \in R_{NP}|\theta_0) = \alpha$ är ett $\alpha$-nivå likformigt starkaste test.*

Notisera att för testerna som uppfyller hypoteserna av Sats 10.1 har vi bara ett element i $\Omega_1$; nämligen, $\theta_1$. Så det finns bara ett meningsfull värde i $\Omega_1$. Det följer att testet som vi får via Sats 10.1 är testet som här den högsta stryka över alla meningsfulla värden i $\Omega_1$ istället för bara stryka minst 0.80 mot alla meningsfulla värden som vi hittade i det föregående underavsnittet.

### 10.2.4 Styrkfunktionen och risk

Hittills har vi utvärderat våra tester genom styrkfunktionen men det är också naturligtvis att betrakta beslutsteori metoder för att utvärdera tester. Precis som vi gjorde med våra punktskattningar kan vi definiera förlustfunktioner på våra tester för att utvärdera deras prestation. Vi ska se att det finns starka kopplingar mellan beslutsteori och styrkfunktion metoderna men beslutsteori perspektivet kommer att låta oss att kvantifiera effekten av ett fel i ett annat sätt.

I ett hypotestest med kritiskt området $R$ har vi vanligtvis två handlingar i handlingsrummet

$$a_0 : \text{ acceptera } H_0 \ (\boldsymbol{x} \notin R),$$
$$a_1 : \text{ acceptera } H_1 \ (\boldsymbol{x} \in R),$$

Notisera att vi kan tänka som testet i formen av en variabel $W(\boldsymbol{X})$ där $W(\boldsymbol{x}) = a_0$ när $\boldsymbol{x} \notin R$ och $W(\boldsymbol{x}) = a_1$ när $\boldsymbol{x} \in R$. Inriktning mot beslutsteori perspektiv specificerar vi nu en förlustfuntion – exempelvis, det generaliserade $(0-1)$-felet

$$L(\theta, a_0) = \begin{cases} 0, & \theta \in \Omega_0, \\ c_{\mathrm{II}}, & \theta \in \Omega_1, \end{cases} \qquad L(\theta, a_1) = \begin{cases} c_{\mathrm{I}}, & \theta \in \Omega_0, \\ 0, & \theta \in \Omega_1, \end{cases}$$

där $c_{\mathrm{I}}$ motsvarar kostnaden för ett Typ-I fel och $c_{\mathrm{II}}$ motsvarar kostnaden för ett Typ-II fel. Nu kan vi beräkna risken för en givet $\theta \in \Omega$ och en viss hypotes $W(\boldsymbol{X})$:

$$
\begin{aligned}
R(\theta, W(\boldsymbol{X}) &= \mathrm{E}[L(\theta, W(\boldsymbol{X})], \\
&= \int_{\boldsymbol{X}} L(\theta, a) f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) d\boldsymbol{x}, \\
&= L(\theta, a_0) \int_{\boldsymbol{X}:W(\boldsymbol{X})=a_0} f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) d\boldsymbol{x} + L(\theta, a_1) \int_{\boldsymbol{X}:W(\boldsymbol{X})=a_1} f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) d\boldsymbol{x}, \\
&= L(\theta, a_0) P(W(\boldsymbol{X}) = a_0|\theta) + L(\theta, a_1) P(W(\boldsymbol{X}) = a_1|\theta), \\
&= L(\theta, a_0) P(\boldsymbol{X} \notin R|\theta) + L(\theta, a_1) P(\boldsymbol{X} \in R|\theta), \\
&= L(\theta, a_0)(1 - \beta(\theta)) + L(\theta, a_1)\beta(\theta).
\end{aligned}
$$

För vårt generaliserad $(0-1)$-fel har vi då att

$$
R(\theta, W(\boldsymbol{X})) = \begin{cases} c_I \beta(\theta) & \theta \in \Omega_0, \\ c_{II}(1 - \beta(\theta)) & \theta \in \Omega_1. \end{cases}
$$

Notisera att $\beta(\theta)$ för $\theta \in \Omega_0$ är sannolikheten av ett Typ-I fel och $1 - \beta(\theta)$ för $\theta \in \Omega_1$ är sannolikheten av ett Typ-II fel. Så vi ser att riskfunktionen och styrkfunktionen är nära sammanlänkande men riskfunktionen inkluderar kostnader av fel kodad i förlustfunktionen.

Vi kan också använder mer sofistikerade förlustfunktioner i sådana analyser. Till exempel kan vi använder förlustfunktioner som ger ett större fel när $\theta$ är längre bort från parametrarna i $\Omega_0$.