

# Föreläsning 9: Hypotestester

Johan Thim (johan.thim@liu.se)

January 15, 2022

Vi har nu studerat metoder för hur man hittar lämpliga skattningar av okända parametrar och även stängt in dessa skattningar i konfidensintervall för att ha kontroll på vad som är rimligt eller ej. Den sista frågan kan man närma sig på lite annorlunda (men egentligen mer naturligt sätt) genom så kallade hypotestester (ibland kallade signifikanstester).

## 1 Hypotestest

Ett hypotesttest i detta sammanhang består av en **nollhypotes**  $H_0$  och en **mothypotes**  $H_1$ . Typiskt är att nollhypotesen är något vi vill motbevisa (och därmed styrka att mothypotesen antagligen gäller). I denna kurs kommer vi oftast begränsa oss till så kallade enkla nollhypoteser och oftast av typen

$$H_0 : \theta = \theta_0.$$

Mothypotesen kan väljas på olika sätt beroende på vad vi vill visa. De vanligaste är av typerna

$$H_1 : \theta \neq \theta_0 \quad \text{eller} \quad H_1 : \theta > \theta_0 \quad \text{eller} \quad H_1 : \theta < \theta_0.$$

Det går att ha betydligt mer komplicerade nollhypoteser (och mothypoteser för den delen). Ett ganska vanligt exempel är  $H_0 : X$  är normalfördelad eller något dylikt. I dessa fall är det svårare att hitta ett lämpligt test. För de enkla typen av nollhypoteser så finns det ganska naturliga teststorheter.

### 1.1 Teststorhet och kritiskt område

För att testa hypotesen behöver vi en teststorhet  $t$  som avgör hur ett stickprov ska behandlas. Denna storhet har analog funktion med de som användes när vi ställde upp konfidensintervall.

Vi låter  $x_1, x_2, \dots, x_n$  vara ett stickprov från en fördelning  $F$  som beror på en okänd parameter  $\theta$ . Motsvarande slumpmässiga stickprov betecknas  $X_1, X_2, \dots, X_n$  i vanlig ordning.



#### Teststorhet/Testvariabel

**Definition.** En funktion  $t : \mathbf{R}^n \rightarrow \mathbf{R}$  given av  $t(x_1, x_2, \dots, x_n)$  kallas **teststorhet** eller **testvariabel** och är en observation av den stokastiska variabeln  $t(X_1, X_2, \dots, X_n)$ .

För att avgöra om vi ska förkasta  $H_0$  väljer vi en **signifikansnivå**  $\alpha$  och bestämmer sedan ett kritiskt område  $C$  som är en delmängd av det område funktionen  $t$  varierar över (en del av värdemängden). Detta område beror på fördelningen  $F$  och den signifikansnivå vi vill utföra hypotestestet på.



## Kritiskt område, signifikansnivå

**Definition.** Det **kritiska området**  $C$  är ett område så att  $H_0$  förkastas om

$$t(x_1, \dots, x_n) \in C.$$

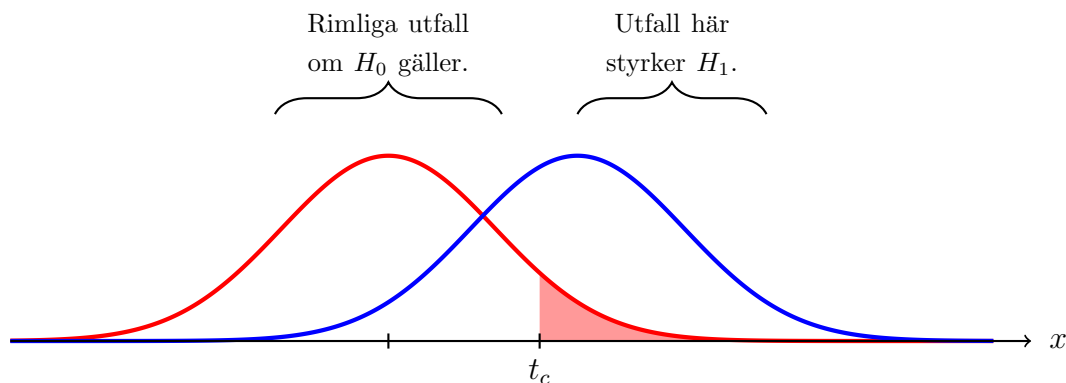
Om  $H_0$  förkastas säger vi att  $H_1$  är styrkt och drar slutsatsen att  $H_1$  gäller. Sannolikheten

$$\alpha = P(t(X_1, \dots, X_n) \in C \mid H_0 \text{ är sann})$$

kallas för testets **signifikansnivå**.

Det kritiska området består alltså av värden som är för extrema för att vara troliga under förutsättningen att nollhypotesen gäller.

Låt oss ställa upp ett hypotestest för väntevärdet för fördelningen  $F$  enligt  $H_0 : \mu = \mu_0$  mot  $H_1 : \mu > \mu_0$ . Vi vill således styrka att det verkliga väntevärdet är större än  $\mu_0$ .

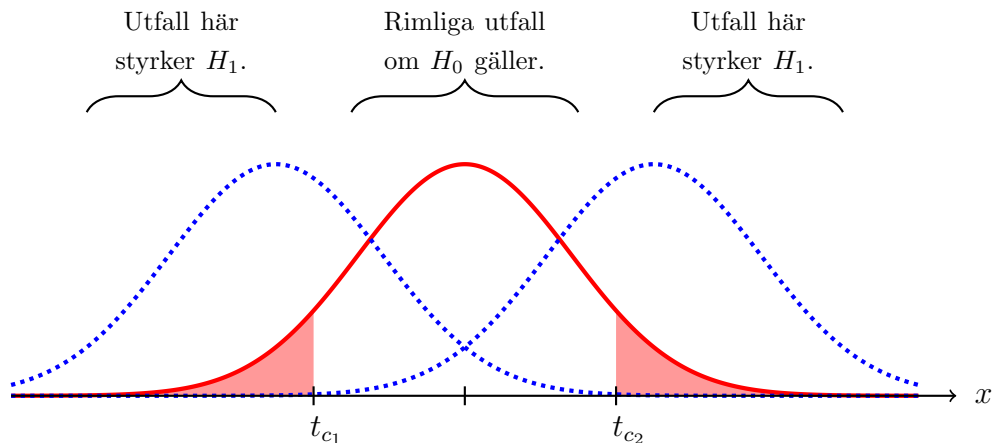


Den röda kurvan är täthetsfunktionen för  $t(X_1, \dots, X_n)$  om  $H_0$  skulle vara sann medan den blå är den verkliga täthetsfunktionen. Vi ser att observerade värden är betydligt rimligare i det kritiska området om den blå fördelningen gäller. Det kritiska området blir således

$$C = \{x \in \mathbf{R} : x > t_c\}.$$

Om  $t > t_c$  så förkastar vi  $H_0$ .

Om vi istället skulle testa  $H_0 : \mu = \mu_0$  mot  $H_1 : \mu \neq \mu_0$ , vad blir skillnaden? Vi vill således i detta läge styrka att det verkliga väntevärdet är något annat än  $\mu_0$  (inte nödvändigtvis att det verkliga väntevärdet är större).



De blå kurvorna är potentiella verkliga fördelningar för  $t(X_1, \dots, X_n)$  medan den röda fortfarande är fördelningen om  $H_0$  skulle vara sann. Det kritiska området blir således

$$C = \{x \in \mathbf{R} : x > t_{c_2} \text{ eller } x < t_{c_1}\}.$$

Om  $t > t_{c_2}$  eller om  $t < t_{c_1}$  så förkastar vi  $H_0$ . När vi vet mer om fördelningen för  $t(X_1, \dots, X_n)$  kan vi under antagandet att  $H_0$  stämmer hitta gränserna explicit.



Att ställa upp  $H_0$  och  $H_1$  ska göras *innan* stickprov observerats. Utgår man från mätdatan för att hitta på sina hypoteser betar man sig bedrägligt.

## 1.2 Styrka, fel och p-värde

Så säg att vi har valt en teststorhet och ett kritiskt område. Vi har då en metod för att förkasta nollhypotesen om teststorheten sticker ut för mycket från vad som är förväntat om nollhypotesen är sann. Om vi då vet vad det verkliga värdet på parametern är, vad blir sannolikheten för att vi kommer att förkasta nollhypotesen? Idealiskt vore den sannolikheten 1 så fort parametern har ett annat värde än vad som angavs in nollhypotesen (dvs  $\theta_0$ ). Detta blir dock svårt att uppfylla, men hur ser sannolikheten ut för att korrekt förkasta  $H_0$  om vi låter  $\theta$  variera? Detta brukar kallas för testets **styrka**.



### Styrka

**Definition.** Vi definierar **styrkefunktionen**  $h(\theta)$  enligt

$$h(\theta) = P(H_0 \text{ förkastas} \mid \theta \text{ är det riktiga värdet}).$$

Sannolikheten  $h(\theta)$  kallas för testets styrka i  $\theta$ .

För ett bra hypotestest bör  $h(\theta)$  vara stor för  $\theta \in H_1$  och  $h(\theta)$  liten för  $\theta \in H_0$ . Notera även att  $h(\theta_0) = \alpha$ . Vi kommer inte att gå in så mycket på styrkan hos tester i denna kurs men det kan vara bra att ha sett begreppet.

Uppenbarligen finns det en risk att vi tar fel beslut. Denna risk kan delas upp i två olika typer.

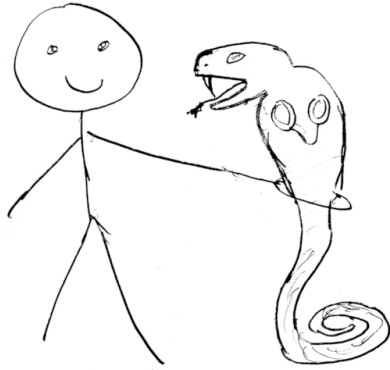


### Fel av typ I och II

**Definition.** Att förkasta  $H_0$  då  $H_0$  är sann kallas **fel av typ I** och har sannolikheten  $\alpha$ . Risken för ett fel av typ I är således signifikansnivån. Att *inte* förkasta  $H_0$  då  $H_0$  är falsk kallas för **fel av typ II**.

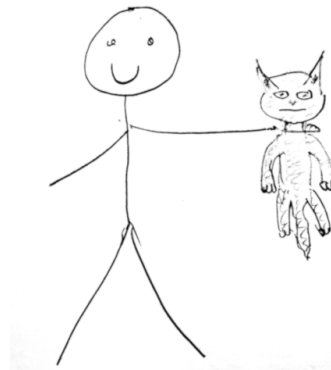
Antag att Bill har lite dålig koll och är ute och letar ormar. När han hittar ett djur så är Bills nollhypotes  $H_0$  : Ormen är giftig.

Fel av typ I



Ormen är inte giftig!

Fel av typ II



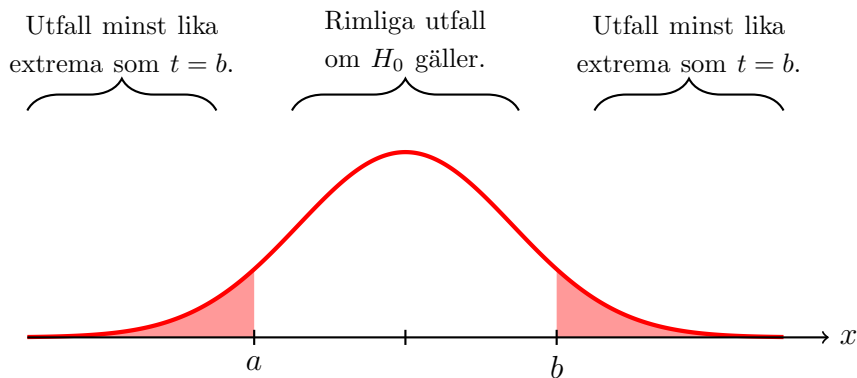
Ormen är giftig!



### *p*-värde

**Definition.** För ett givet stickprov kan man för ett signifikanstest beräkna ett *p*-värde. Denna sannolikhet är den lägsta signifikansnivån på vilken vi skulle förkasta  $H_0$ . Med andra ord är *p* sannolikheten att vi får ett minst lika extremt utfall som det givna stickprovet med antagandet att  $H_0$  är sann.

Låt oss testa  $H_0 : \theta = \theta_0$  mot  $H_1 : \theta \neq \theta_0$ . Om vi utifrån stickprovet beräknar teststorheten  $t(x_1, \dots, x_n) = b$  så behöver vi alltså karakterisera alla utfall som är minst lika extrema om  $H_0$  gäller. Nu blir vi beroende av hur fördelningen ser ut. Låt oss anta något symmetriskt.



Så *p*-värdet kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$p = P(t(X_1, \dots, X_n) \leq a) + P(t(X_1, \dots, X_n) \geq b) = 2P(t(X_1, \dots, X_n) \geq b),$$

där *a* måste väljas så vi har samma sannolikhetsmassa i båda "svansarna." Om fördelningen har en riktig skum uppsyn då? Ja, då blir det svårt. En variation vi kan hantera är om mothypotesen är av typen  $H_1 : \theta > \theta_0$  (till exempel) då vi endast har

$$p = P(t(X_1, \dots, X_n) \geq b)$$

eftersom utfall i vänstra svansen nu inte längre räknas som extrema. Utseendet på mothypotesen är alltså fundamentalt.



Märk väl att  $p$ -värdet inte säger någonting om huruvida  $H_0$  är sann eller ej givet observationen av  $t$ . Det vi har är sannolikheten för ett lika extremt utfall *givet* att  $H_0$  gäller. *Inte* tvärtom!

Alla principfigurer ovan har varit små söta symmetriska och kontinuerliga historier. Hur blir det vid andra typer av fördelningar?

## 2 Hypotestest för Binomialfördelning

Vi undersöker situationen med ett belysande exempel.



### Exempel

Ett mynt kastas (oberoende) 30 gånger och vid 10 av dessa blir det en krona. Kan vi förkasta hypotesen att myntet är ärligt med signifikansnivå 5%?

**Lösning.** Vi vill testa om myntet är ärligt, så vi börjar med att ställa upp en modell. Låt  $X$  vara antalet krona vid 30 kast. Då är  $X \sim \text{Bin}(n, p)$  där  $n = 30$  och  $p =$  sannolikheten för krona är okänd. Så en rimlig nollhypotes ges av

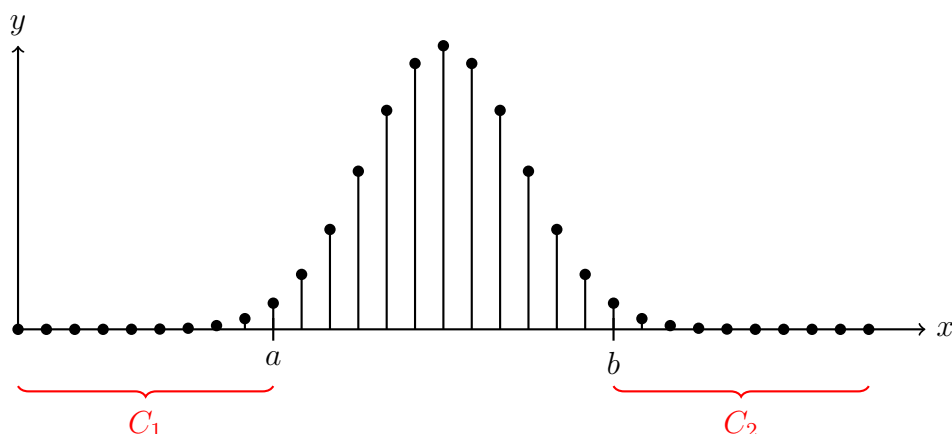
$$H_0 : p = \frac{1}{2}$$

och innan experimentet vet vi inte om mothypotesen bör vara  $p < 1/2$  eller  $p > 1/2$ , så vi tar det säkra före det osäkra och väljer att testa mot

$$H_1 : p \neq \frac{1}{2}.$$

Givet att  $H_0$  är sann så förväntar vi oss frekvensen  $30 \cdot 0.5 = 15$  utfall som är krona. Är 10 signifikant mindre? Vi ställer upp det kritiska området:

$$C = \{x \in \mathbf{Z} : 0 \leq x \leq a \text{ eller } b \leq x \leq n\}$$



Hur hittar vi  $a$  och  $b$ ? Vi får helt enkelt testa oss fram (och använda tabeller). Eftersom

$$p(x) = \binom{30}{x} \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{30-x}$$

kan vi beräkna att

$$\sum_{x=0}^9 p(x) = 0.0214 \quad \text{och} \quad \sum_{x=0}^{10} p(x) = 0.0494$$

samt (känt redan pga symmetri då  $p = 0.5$  men för fullständighetens skull):

$$\sum_{x=21}^{30} p(x) = 0.0214 \quad \text{och} \quad \sum_{x=20}^{30} p(x) = 0.0494.$$

Vi väljer  $a = 9$  och  $b = 21$ . Då gäller att

$$\begin{aligned} P(X \in C | H_0) &= P(X \in C_1 | H_0) + P(X \in C_2 | H_0) \\ &= P(X \leq a) + P(X \geq b) = 0.0214 + 0.0214 = 0.0428 < 0.05. \end{aligned}$$

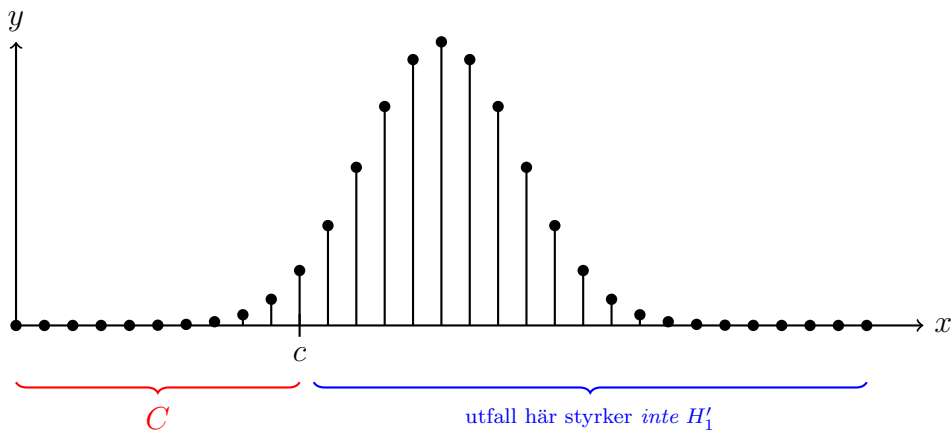
Detta är det största kritiska område vi kan få för att hålla signifikansnivån. Observera att vi alltså inte kan träffa  $\alpha = 0.05$  exakt. Detta är typiskt vid diskreta fördelningar.

Eftersom  $x = 10 \notin C$  kan vi inte dra någon slutsats, utan myntet kan mycket väl vara ärligt. Vi kan således *inte* förkasta  $H_0$  (vilket inte på något sätt betyder att  $H_0$  är sann).

Antag att vi istället vill testa mothypotesen  $H'_1$  att myntet ger färre krona än klave. Vi har då

$$H'_1 : p < \frac{1}{2}.$$

Hur ser det kritiska området  $C$  ut?



Eftersom

$$\sum_{x=0}^{10} p(x) = 0.0494 \quad \text{och} \quad \sum_{x=0}^{11} p(x) = 0.1002$$

så ser vi att  $c = 10$  är nödvändigt. Därmed blir

$$C = \{x \in \mathbf{Z} : 0 \leq x \leq 10\}$$

och vår observation  $x = 10 \in C$ . Alltså kan vi förkasta  $H_0$  och anse att  $H'_1$  är styrkt. Slutsats?

### 3 Normalapproximation – Generellt

När vi approximerar med normalfördelningen är tillvägagångssättet nästan alltid det samma. Vi har en punktskattning  $\hat{\theta}$  där  $\hat{\Theta} \stackrel{\text{appr.}}{\sim} N(\theta, D^2)$  och vi vill testa nollhypotesen  $H_0 : \theta = \theta_0$ . Som teststorhet använder vi då oftast

$$Z = \frac{\hat{\Theta} - \theta_0}{D} \quad \text{eller} \quad Z = \frac{\hat{\Theta} - \theta_0}{d}.$$

Den senare teststorheten då vi inte känner  $D$  exakt utan skattar med  $d$ . Vi förutsätter att  $d$  är en vettig skattning av  $D$  då  $H_0$  är sann. Notera att i båda fallen kommer  $Z \stackrel{\text{appr.}}{\sim} N(0, 1)$  om  $H_0$  är sann. Vi använder alltså ingen  $t$ -fördelning här (det finns inget som säger att det skulle bli bättre i det generella fallet).

Hur det kritiska området ser ut beror på hur vi ställer upp mothypotesen. Om  $H_1 : \theta \neq \theta_0$  får  $C$  utseendet  $]-\infty, -a[ \cup ]a, \infty[$ . Är mothypotesen enkelsidig blir det bara ett av intervallen (med annan parameter  $a$ ). Talet  $a$  hittar vi i normalfördelningstabell.

### 4 Test för skillnad i andel

En mycket vanlig situation är att vi vill undersöka om det föreligger någon skillnad i andel mellan två grupper. Antag att vi har  $x_1$  som observation av  $X_1 \sim \text{Bin}(n_1, p_1)$  och  $x_2$  som observation av  $X_2 \sim \text{Bin}(n_2, p_2)$  (vi antar oberoende).

Vi är intresserade av att testa hypotesen  $H_0 : p_1 = p_2$  mot till exempel  $H_1 : p_1 \neq p_2$ . Om  $H_0$  är sann så är en lämplig skattning av  $p = p_1 = p_2$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

Faktum är att detta är ML-skattningen (om  $H_0$  är sann) och därmed har den bra egenskaper såsom konsistens. Vad gäller fördelningen för  $\hat{P}$  blir den värre (vad händer om man summerar binomialfördelningar?). Men, om  $n_1$  och  $n_2$  är ganska stora och  $p$  inte är allt för nära ändpunkterna i  $[0, 1]$ , så kanske vi kan normalapproximera? Vi har redan gjort detta (se konfidensintervall för  $p_1 - p_2$ ), men för fullständighetens skull låt oss repetera. Om  $H_0$  är sann gäller att

$$E(\hat{P}) = \frac{n_1 p + n_2 p}{n_1 + n_2} = p$$

och

$$V(\hat{P}) = \frac{n_1 p(1-p) + n_2 p(1-p)}{(n_1 + n_2)^2} \rightarrow 0,$$

då  $n_1 + n_2 \rightarrow \infty$ , så skattningen av  $p$  är väntevärdesriktig och konsistent. För att testa  $H_0$  använder vi  $\hat{P}_1 - \hat{P}_2$ , och om  $H_0$  är sann så gäller att

$$\hat{P}_1 - \hat{P}_2 \stackrel{\text{appr.}}{\sim} N\left(0, \hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

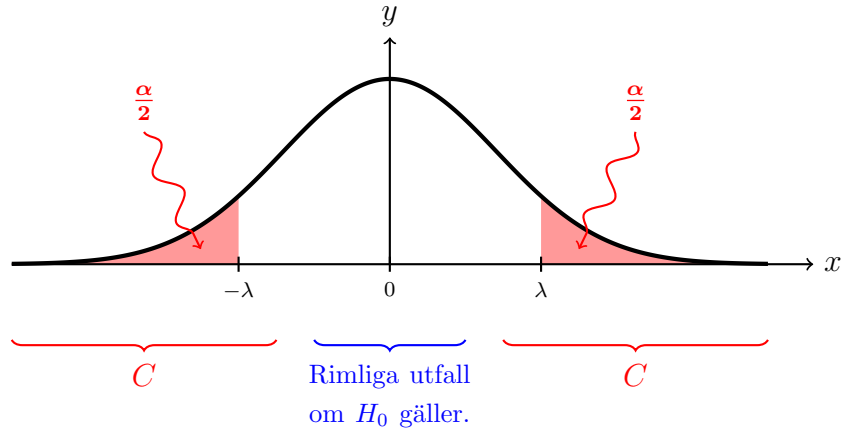
Eftersom vi inte känner  $p$  exakt använder vi skattningen  $\hat{p}$  ovan i uttrycket för variansen (eller vi ersätter standardavvikelsen med medelfelet). Vi kan även gå över i standardiserad form så vi känner igen oss:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \stackrel{\text{appr.}}{\sim} N(0, 1).$$

Om  $H_1 : p_1 \neq p_2$  så ges det kritiska området av

$$C = \{z \in \mathbf{R} : |z| > \lambda\}$$

för något lämpligt  $\lambda = \Phi^{-1}(1 - \alpha/2)$  vi finner ur tabell (eller MATLAB).



### Exempel

Två opinionsinstitut Analysera Mera AB och StickProvarna AB undersöker om befolkningen tycker att sommaren varit för varm. AM frågar 500 personer och andelen  $p_1 = 0.7$  (350 st) håller med. SP frågar 400 personer och  $p_2 = 0.8$  (320 stycken) håller med. Undersök om det finns någon signifikant skillnad mellan resultaten på signifikansnivån 5% (approximativt).

**Lösning.** Låt  $H_0 : p_1 = p_2 = p$  och  $H_1 : p_1 \neq p_2$ . Om  $H_0$  är sann väljer vi skattningen

$$\hat{p} = (350 + 320)/(500 + 400) = 0.744.$$

Med beteckningarna ovan gäller då (om  $H_0$  är sann) att

$$Z = \frac{\widehat{P}_1 - \widehat{P}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{500} + \frac{1}{400}\right)}} = \frac{\widehat{P}_1 - \widehat{P}_2}{0.0293} \underset{\text{appr.}}{\sim} N(0, 1).$$

Det är rimligt att approximera både  $\widehat{P}_1$  och  $\widehat{P}_2$  med normalfördelning eftersom både  $500 \cdot 0.7 \cdot 0.3 \geq 10$  och  $400 \cdot 0.8 \cdot 0.2 \geq 10$ . Vi hittar det kritiska området

$$C = \{z \in \mathbf{R} : |z| > \lambda\}$$

där  $\lambda = \Phi^{-1}(0.975) = 1.96$ . Således ska – om  $H_0$  är sann –

$$\left| \frac{\widehat{p}_1 - \widehat{p}_2}{0.0293} \right| > 1.96 \quad \Leftrightarrow \quad |\widehat{p}_1 - \widehat{p}_2| > 1.96 \cdot 0.0293 = 0.0573$$

för att vi ska förkasta  $H_0$ . Med  $\widehat{p}_1 = 0.7$  och  $\widehat{p}_2 = 0.8$  ser vi att  $0.1 > 0.0573$ , så vi förkastar  $H_0$ . Det är troligen en skillnad i resultaten.

Ett alternativ är att ställa upp konfidensintervallet  $I_{p_1 - p_2}$  för  $p_1 - p_2$  och sedan testa hypotesen genom att undersöka om  $0 \in I_{p_1 - p_2}$ . Skulle det vara så att 0:an ingår kan vi inte förkasta  $H_0$ . Ligger intervallet helt på ena sidan 0 däremot så förkastar vi  $H_0$ . Detta test är helt ekvivalent eftersom vi nyttjar samma testvariabel.



## 5 Poissonapproximation

Som bekant kan man även approximera binomialfördelning med Poissonfördelning om  $n \geq 10$  och  $p \leq 0.1$ . Detta kan vara nödvändigt då  $p$  ligger nära 0 eller 1 så normalapproximation inte fungerar bra. Vi betraktar ett exempel.



### Exempel

En leverantör av laboratorieutrustning hävdar att deras pipetter bara behöver kalibreras en gång per år och att risken för att en pipett faller utanför toleransnivån innan dess är 0.5% (vid normal användning). Laborarieansvarig Laura (för ett stort laboratorie) tycker inte att det stämmer och har ett år efter inköpet och kontinuerligt användande av 1000 stycken behövt kalibrera om 11 st. Testa hypotesen att felrisken är 0.5% mot att den är högre på signifikansnivån 1% (approximativt).

**Lösning.** Den stokastiska variabeln  $X$  är antalet av de 1000 pipetterna som behövs kalibreras i förtid. Om vi antar att händelserna är oberoende (är det rimligt?) så är  $X \sim \text{Bin}(1000, p)$  där  $p$  är felrisken. Låt  $H_0 : p = 0.005$  och  $H_1 : p > 0.005$ . Vi kan använda  $\hat{P} = \frac{X}{1000}$ , men enklare är att direkt nyttja  $X$ . Om  $H_0$  är sann så gäller att

$$X \stackrel{\text{appr.}}{\sim} \text{Po}(1000 \cdot 0.005) = \text{Po}(5).$$

Det kritiska området väljs som

$$C = \{z \in \mathbf{Z} : z > k\}$$

för något  $k \in \mathbf{Z}$ . Vi vill att

$$P(X \in C | H_0) \leq 0.01$$

och i tabell (eller med  $\mathbf{k} = \text{poissinv}(0.99, 5)$  i MATLAB, vilket ger det minsta heltalet  $k$  så att  $P(X \leq k) \geq 0.99$ ) finner vi att  $k = 11$ . Alltså gäller

$$P(X > 11 | H_0) < 0.01 \quad (\text{exakt värde: } 0.0055),$$

och Lauras observation  $x = 11$  är alltså *inte* signifikant. Vi kan inte förkasta  $H_0$  och säga att leverantören har fel.



### Exempel

Laura är inte nöjd och kräver att examensarbetaren Audrey ska göra om hypotestestet och använda normalapproximation *som folk*. Motivera varför det inte är bra men utför testet. Undersök också hur hypotestestet blir om man inte approximerar för att hjälpa den stackars examensarbetaren att motivera.

**Lösning.** Vid normalapproximation kräver vi att  $np(1-p) \geq 10$  och om vi väljer att skatta  $p$  med  $\hat{p} = 10/1000 = 0.01$  hamnar vi precis kring den gränsen så osäkerheten är stor. Använder vi leverantörens  $p = 0.005$  blir det betydligt under. Alltså inget att rekommendera. Men om vi envisas så skulle

$$X \stackrel{\text{appr.}}{\sim} N(1000p, 1000p(1-p)) \quad \text{som dålig approximation.}$$

Om vi antar att  $H_0$  är sann skulle då

$$Z = \frac{X - 1000 \cdot 0.005}{\sqrt{1000 \cdot 0.005 \cdot (1 - 0.005)}} \stackrel{\text{appr.}}{\sim} N(0, 1),$$

återigen som en tveksam approximation. Kritiskt område ges av

$$0.01 = P(Z > \lambda) = 1 - \Phi(\lambda) \Leftrightarrow \lambda = \Phi^{-1}(0.99) = 2.3263$$

så

$$\frac{X - 5}{\sqrt{4.975}} > 2.3263 \Leftrightarrow X > 10.1888$$

och vi skulle därför låta  $C$  ges av  $X \geq 11$ , varvid resultatet  $x = 11$  skulle verka signifikant.

Vi kan ställa upp ett exakt test genom att låta  $H_1 : p > 0.005$  och välja

$$C = \{x \in \mathbf{Z} : x > k\}$$

för något  $k \in \mathbf{Z}$ . Precis som med Poissonapproximationen hittar vi  $k$  genom att i MATLAB använda `k = binoinv(0.99, 1000, 0.005)` vilket resulterar i  $k = 11$ . Alltså samma gräns som vi fick med Poissonapproximationen. Exakt värde här blir  $P(X > k) = 0.0053$ .

## 6 Hypotestest för väntevärde vid normalfördelning

Låt oss undersöka den generella situationen vid ett stickprov  $X_1, X_2, \dots, X_n$  från en normalfördelning  $N(\mu, \sigma)$ . Vi kan tänka oss att

$$X_i = \mu + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma), \quad i = 1, 2, \dots, n,$$

där  $\epsilon_i$  är oberoende. Notera att samtliga variabler har samma varians. Som föregående avsnitt visade är det skillnad på när vi känner variansen exakt och när den behöver skattas.

Vi börjar med att testa

$$H_0 : \mu = \mu_0 \quad \text{mot} \quad H_1 : \mu \neq \mu_0.$$

Givetvis kan man vilja testa mot  $H_1' : \mu > \mu_0$  eller  $H_1'' : \mu < \mu_0$  också, vi kommer ta upp något sådant exempel senare.

### 6.1 Känd varians

Eftersom

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

kan vi när  $\sigma$  är känd direkt använda  $\bar{X}$  som teststorhet. Men för att göra det hela systematiskt och analogt med fallet då  $\sigma$  inte är känd skapar vi en testvariabel

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

där fördelningen gäller under förutsättning att  $H_0$  är sann. Vad vi egentligen gör är att vi utnyttjar att  $\bar{X}$  är en skattning (konsistent och väntevärdesriktig) av (det okända) väntevärdet  $\mu$ . Testet går ut på att se om det uppmätta värdet på skattningen sticker ut så mycket från vad som är förväntat att det gör  $H_0$  orimlig.

Det kritiska området  $C$  ges av

$$P(Z \in C | H_0) = \alpha$$

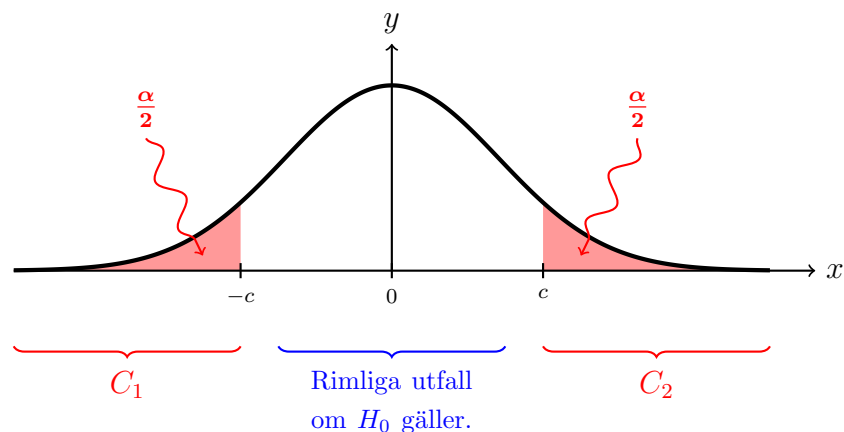
där vi av symmetriskäl (eftersom  $Z \sim N(0, 1)$ ) kan – för något  $c > 0$  – uttrycka  $C$  enligt

$$C = \{z \in \mathbf{R} : |z| > c\} = \{z \in \mathbf{R} : z > c \text{ eller } z < -c\}$$

Vi noterar att  $C$  består av två delar  $C_1$  och  $C_2$  där talen i  $C_1$  är negativa och talen i  $C_2$  är positiva. Återigen, av symmetriskäl måste

$$P(Z \in C_1) = P(Z \in C_2) = \frac{\alpha}{2}.$$

Gränsen hittar vi i tabell genom att leta reda på ett tal  $c = \Phi^{-1}(1 - \alpha/2)$  (sitter du med MATLAB kan du använda `c = -norminv(alpha/2)`).



### 6.1.1 Approximativt test via CGS

Som vi såg på förra föreläsningen kan man använda approximationer för att utföra hypotestest. Om vi i vårt fall inte vet att  $X_i$  är normalfördelad kan vi ändå via centrala gränsvärdessatsen säga att

$$\bar{X} \stackrel{\text{appr.}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

om  $n \geq 30$  (lite beroende på hur skev fördelningen för  $X_i$  är). Som teststorhet använder vi sedan

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

Notera att vi ersätter  $\sigma$  med  $s$  utan att förändra fördelningen (eftersom vi redan håller på med approximationer vet vi inte om det blir bättre med  $t$ -fördelningen). Faktum är att vi kan göra detta även om variablerna är lite beroende. Det finns flera varianter av CGS som kan hantera lite olika situationer.

## 6.2 Okänd varians

Om vi inte känner till  $\sigma$  så kan vi inte direkt använda  $\bar{X}$  som teststorhet och inte heller  $Z$  från föregående stycke fungerar utan problem (vad ska vi göra med den okända storheten  $\sigma$ ?). Visserligen kan vi se det approximativt via CGS, men vad som vore bättre är att ersätta  $\sigma^2$  med stickprovsvariansen  $s^2$ . Som vi sett tidigare leder det till  $t$ -fördelningen. Så, då gäller att

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$$

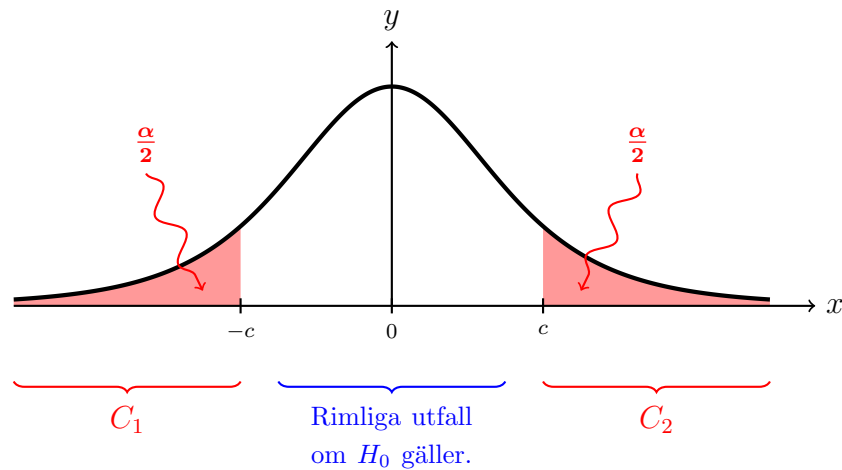
om  $H_0$  är sann. Helt analogt med föregående situation erhåller vi nu

$$C = \{t \in \mathbf{R} : t > c \text{ eller } t < -c\}$$

där

$$P(T > c) = P(T < -c) = \frac{\alpha}{2}.$$

Gränsen hittar vi i tabell genom att leta reda på ett tal  $c = F_T^{-1}(1 - \alpha/2)$  (sitter du med MATLAB kan du använda  $c = -\text{tinv}(\alpha/2)$ ).



## 7 Hypotestester och konfidensintervall

Den observante läsaren har nog redan reflekterat över att det vi ägnat oss åt är ganska snarlikt de föregående föreläsningarna om konfidensintervall. Vi ställer upp liknande storheter (ja, identiska för det mesta) men istället för att stänga in något okänt i ett intervall så testar vi skattningen mot ett kritiskt område.

Ett annat sätt att testa hypoteserna på är att ställa upp konfidensintervall och sedan testa om intervallet täcker nollhypotesen eller ej.



### Exempel

I ett husvagnsdrivet laboratorie kokar Janne ihop den suspekta kemikalien  $C_{10}H_{15}N$ . Varje vecka startar han med samma mängd utgångsmaterial och följer samma procedur. Janne har fått en ny köpare och har hävdats att han kan producera 500 gram i veckan. För att inte riskera problem med hälsan vill Janne testa hypotesen  $H_0 : \mu = 500$  mot  $H_1 : \mu > 500$  på signifikansnivån 1%. Under 16 veckor producerar han i snitt 525 gram med stickprovsstandardavvikelsen 30 gram.

**Lösning.** Vi antar normalfördelning och ställer upp ett enkelsidigt konfidensintervall för väntevärdet  $\mu$ . Låt

$$T = \frac{\bar{X} - \mu}{S/\sqrt{16}} \sim t(15).$$

Då gäller att

$$P(T < t) = 0.99$$

om  $t = 2.6025$  (ur tabell). Eftersom

$$T < t \Leftrightarrow \frac{\bar{X} - \mu}{S/4} < t \Leftrightarrow \frac{St}{4} > \bar{X} - \mu \Leftrightarrow \mu > \bar{X} - \frac{St}{4}.$$

så erhåller vi konfidensintervallet

$$I_\mu = \left( \bar{x} - \frac{st}{4}, \infty \right) = (505.48, \infty).$$

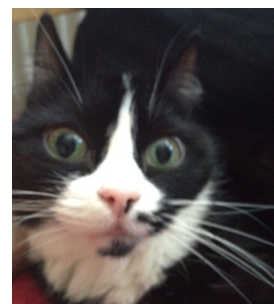
Om  $H_0$  är sann så kommer  $\mu = 500 \in I_\mu$  med sannolikheten 99%. Eftersom detta inte är sant kan vi förkasta  $H_0$ . Ska Janne sitta lugnt i båten att han inte lovat för mycket? Mäter Jannes test rätt sak?

## 8 Generellt om hypotestester

Innan vi avslutar med en diskussion och tester vid flera stickprov tar vi och summerar lite att tänka på.

- (i) Formulera hypoteser *innan* du tittar på datan. Att du vill göra ett enkelsidigt test ska inte bero på hur datan ser ut. Av den anledningen är den vanligaste typen av tester två-sidiga.
- (ii) Men vissa situationer är alltid enkelsida på grund av konstruktion. Vi kommer se det i samband med regressionsanalysen där vi till exempel vet att varians minskar med fler förklaringsvariabler.
- (iii) Kom ihåg när hypotestestet ställs upp att det är *mothypotesen* vi vill styrka.
- (iv) Var *mycket* försiktig med tolkning av resultaten.

Att man inte förkastar  $H_0$  betyder inte att  $H_0$  gäller. Ett klassiskt exempel handlar om fyrbenta djur: låt  $H_0$  : djuret har fyra ben och  $H_1$  : djuret har inte fyra ben. Vi vill undersöka om en observation är en häst och testar  $H_0$  mot  $H_1$ . Bara för att vi inte kan förkasta  $H_0$  när djuret är en katt betyder det inte att det är en häst, eller hur? Kanske ett urartat exempel, det kan vara betydligt mer diffust att läsa av resultaten rätt i andra fall.



- (v) Signifikansnivån kan också vara missvisande. Vid stora stickprov kan man ofta se en skillnad och förkasta  $H_0$  även om skillnaden kanske inte spelar någon större roll i praktiska fall.

## 9 Flera stickprov

Vi kan givetvis betrakta flera stickprov samtidigt. Ofta är man intresserad av att testa om de har samma väntevärde och/eller samma varians. Men vi kan ställa upp test för linjärkombinationer av väntevärdena direkt.

Låt  $X_1, X_2, \dots, X_m$  och  $Y_1, Y_2, \dots, Y_n$  vara stickprov från  $N(\mu_X, \sigma_X)$  respektive  $N(\mu_Y, \sigma_Y)$ . Vi kan ställa upp hypotestester för linjärkombinationen  $c_1\mu_X + c_2\mu_Y$ . Om varianserna är kända kan vi direkt använda att

$$Z = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_X + c_2\mu_Y)}{\sqrt{c_1^2\sigma_X^2/m + c_2^2\sigma_Y^2/n}} \sim N(0, 1),$$

alltså precis samma variabel vi såg när vi tog fram konfidensintervall för  $c_1\mu_X + c_2\mu_Y$ .

## 9.1 $\sigma_X = \sigma_Y = \sigma$ okänd

Helt analogt med motsvarande situation när vi tog fram konfidensintervall använder vi att

$$T = \frac{c_1\bar{X} + c_2\bar{Y} - (c_1\mu_X + c_2\mu_Y)}{S\sqrt{c_1^2/m + c_2^2/n}} \sim t(m+n-2),$$

där  $S^2$  är den sammanvägda variansskattningen. Vi betraktar ett exempel.



### Exempel

Janne har fått konkurrens av den före detta lärlingen Rossana som använder samma metod. Under 9 veckor producerar hon i snitt 600 gram med en stickprovsstandardavvikelse på 50 gram. Testa på signifikansnivån 1% hypotesen  $H_0 : \mu_1 = \mu_2$  mot  $H_1 : \mu_1 < \mu_2$  med antagandet att variansen är densamma, där  $\mu_1$  är Jannes förväntade värde och  $\mu_2$  är Rossanas. Borde köparen byta leverantör?

**Lösning.** Vi formulerar om enligt  $H_0 : \mu_2 - \mu_1 = 0$  mot  $H_1 : \mu_2 - \mu_1 > 0$ . Om  $H_0$  är sann så gäller att

$$T = \frac{\bar{Y} - \bar{X}}{S\sqrt{1/9 + 1/16}} = \frac{\bar{Y} - \bar{X}}{0.4167 S} \sim t(23),$$

och det kritiska området blir

$$C = \{t \in \mathbf{R} : t > 2.4999\}$$

eftersom  $P(T < 2.4999) = 0.99$ . Med uppmätta siffrorna blir

$$t = \frac{600 - 525}{0.4167 \cdot s_p} = \frac{75}{0.4167 \cdot 38.16} = 4.7161,$$

där

$$s_p^2 = \frac{15 s_1^2 + 8 s_2^2}{23} = 38.16^2$$

är den sammanvägda variansskattningen. Eftersom  $4.7161 \in C$  förkastar vi  $H_0$ . Blir det samma resultat om vi testar mot  $H_1 : \mu_1 \neq \mu_2$ ? (svar: ja, förkasta  $H_0$ . Vad ändras?)

Vi kan givetvis ta fram ett konfidensintervall  $I_{\mu_2 - \mu_1}$  och testa nollhypotesen genom att undersöka om  $0 \in I_{\mu_2 - \mu_1}$  också (förkasta  $H_0$  om  $0 \notin I_{\mu_2 - \mu_1}$ ).

## 10 (★) Hypotestest för Poissonfördelning

Övriga diskreta fördelningar kan givetvis hanteras analogt med binomialexemplet i föregående avsnitt och med datorkraft är det inte större problem att räkna exakt i väldigt många fall. Men som vi kommer ihåg från tidigare kurser går det även att approximera flera diskreta fördelningar med normalfördelning om vissa förutsättningar är uppfyllda. Låt oss studera ett exempel med Poissonfördelning på två sätt.



### Exempel

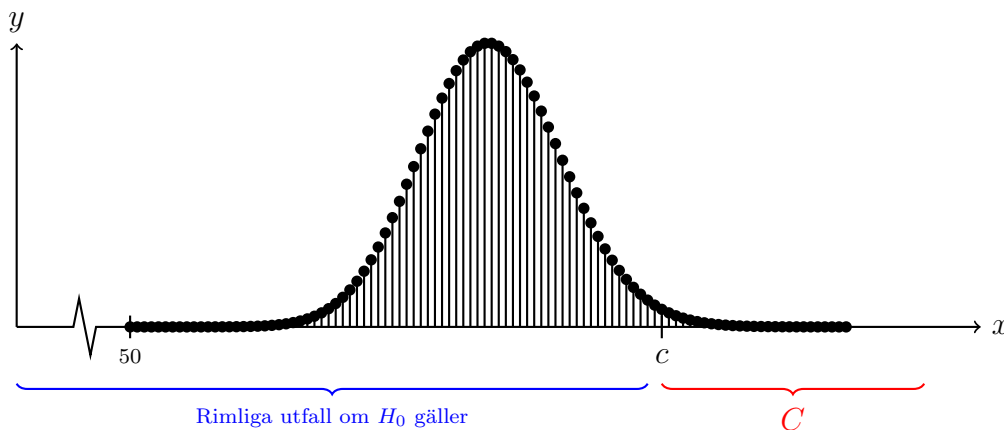
Antalet datapaketer till en server kan betraktas som en Poissonprocess  $X(t)$  med en okänd intensitet  $\lambda$ . För att kunna hantera överbelastning har man ett varningssystem som varnar om antalet paket överstiger en gräns  $N$  på två tidsenheter. Varningen sker alltså om intensiteten är större än väntat. Antag att  $\lambda = 50$  (enhet: tusen paket). Det är dyrt att avbryta servicen så man vill högst tillåta felaktig varning med 1% risk.

Hitta gränsen  $N$  och avgör om man bör varna om  $x = 120$  vid en mätning. Vad skulle  $p$ -värdet bli om  $x = 130$ ?

**Lösning.** Det förväntade antalet paket är  $\mu = E(X(t)) = \lambda t$ , så om  $\lambda = 50$  förväntar vi oss  $\mu = 50 \cdot 2 = 100$  (tusent) paket. Låt

$$H_0 : \mu = 100 \quad \text{och} \quad H_1 : \mu > 100.$$

Vi söker det kritiska området  $C$ . En figur kan vara bra.



Låt  $p(k)$ ,  $k = 0, 1, 2, \dots$ , vara sannolikhetsfunktionen för en  $Po(100)$ -fördelad variabel. Ur tabell (eller med hjälp av matlab och funktionerna `poisspdf` eller `poisscdf`) kan vi finna att

$$\sum_{k=124}^{\infty} p(k) = 1 - \sum_{k=0}^{123} p(k) = 0.0112 \quad \text{och} \quad \sum_{k=125}^{\infty} p(k) = 0.0088.$$

Således blir det kritiska området

$$C = \{k \in \mathbf{Z} : k \geq 125\}.$$

Eftersom observationen  $x = 120 \notin C$  så kan vi inte förkasta  $H_0$ . Vi bör inte varna.

Vi beräknar  $p$ -värdet vid observationen  $x = 130$  genom

$$p = P(X \geq 130 | H_0) = \sum_{k=130}^{\infty} p(k) = \text{tabell} = 0.0023.$$

Vi summerar alltså sannolikheterna för alla utfall som är minst lika extrema som  $x = 130$ .

Om vi stirrar lite på plotten ovan så ser den tämligen normalfördelad ut, eller hur? Det är ingen slump. Om  $X \sim \text{Po}(\mu)$  med  $\mu \geq 15$  så är  $X \stackrel{\text{appr.}}{\sim} N(\mu, \sqrt{\mu})$  (variansen är  $\mu$ ). Vi kan använda detta för att hitta en approximativ gräns  $N$ . Låt  $X \sim \text{Po}(100)$ . Då gäller att

$$0.01 = P(X \geq N) = 1 - P(X < N) = 1 - P\left(\frac{X - 100}{\sqrt{100}} < \frac{N - 100}{\sqrt{100}}\right) = 1 - \Phi\left(\frac{N - 100}{10}\right).$$

Således är

$$\begin{aligned} 0.01 = 1 - \Phi\left(\frac{N - 100}{10}\right) &\Leftrightarrow 0.99 = \Phi\left(\frac{N - 100}{10}\right) \\ &\Leftrightarrow 2.3263 = \frac{N - 100}{10} \\ &\Leftrightarrow N = 23.263 + 100 = 123.263. \end{aligned}$$

Eftersom  $N$  måste vara ett heltal väljer vi  $N = 124$ . Även med halvstegskorrigering hamnar vi inte på det exakta värdet, men det är tillräckligt nära för de flesta ändamål. Vi kan även återskapa kalkylen för  $p$ -värdet vid  $x = 130$  enligt

$$p \approx 1 - \Phi\left(\frac{130 - 100}{10}\right) = 0.0013.$$

## 11 (★)Plåtfabriken

I detta avsnitt följer en sekvens av exempel som kanske illustrerar en del av föregående material.



### Exempel

I en fabrik med mångårig erfarenhet tillverkar man material av en viss tjocklek. Man mäter med jämna mellanrum tjockleken på 9 nytillverkade material och testar om medelvärdet uppfyller  $|\bar{x} - 5.0| > 0.05$ . Om så är fallet stoppas tillverkningen och tekniker får gå igenom maskineriet. Ansvarig för metodutvecklingen vet av erfarenhet att  $\sigma = 0.1$ . Om vi antar normalfördelning, vad är bästa signifikansnivån för testet om  $H_0 : \mu = 5$  testas med mothypotesen  $H_1 : \mu \neq 5$ ?

**Lösning.** Vi antar att  $X_i \sim N(\mu, \sigma) = N(\mu, 0.1)$ ,  $i = 1, 2, \dots, 9$ , är oberoende. För att testa  $H_0$  ställer vi upp teststorheten

$$Z = \frac{\bar{X} - 5.0}{0.1/\sqrt{9}}.$$

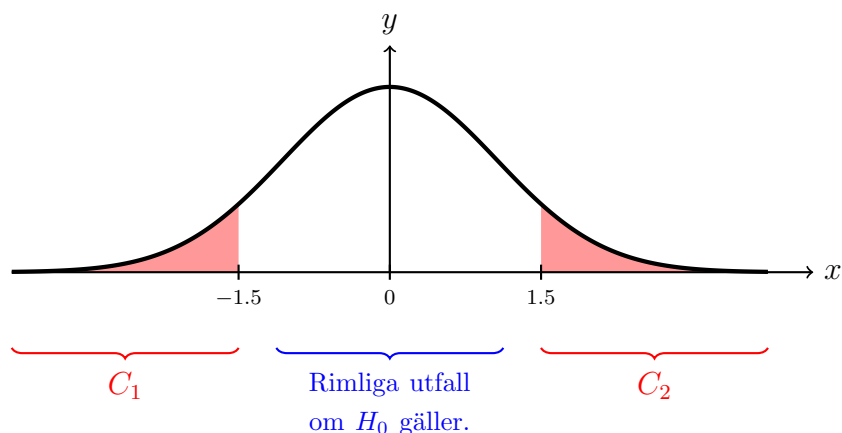
Om  $H_0$  är sann så är  $Z \sim N(0, 1)$ . Om vi jämför med fabriken test så ser vi att

$$\left|\frac{0.1}{3} Z\right| > 0.05 \Leftrightarrow |Z| > 1.5$$

ger det kritiska området

$$C = \{z \in \mathbf{R} : |z| > 1.5\}.$$





Så signifikansnivån  $\alpha$  kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$p = P(Z \leq -1.5) + P(Z \geq 1.5) = 2P(Z \leq -1.5) \\ = 2\Phi(-1.5) = 2(1 - \Phi(1.5)) = 0.1336.$$

Den bästa signifikansnivån vi kan välja är alltså  $\alpha = 0.1336$ .



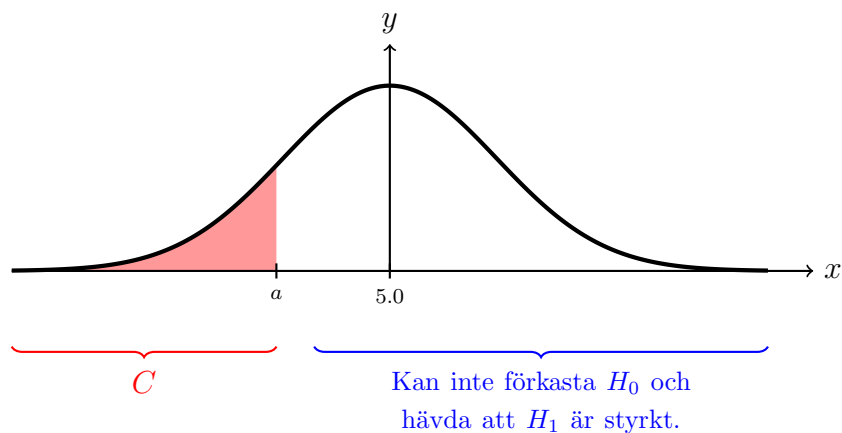
### Exempel

Fabriken har fått en ny beställare som inte har något problem om materialet blir tjockare. Ansvarig tänker lite snabbt och ställer upp ett test med samma signifikansnivå för att endast testa att materialet inte blir för tunt. Hur ser testet ut nu och varför är detta antagligen inte vad man vill göra?

**Lösning.** Vi har fortfarande  $H_0 : \mu = 5$  men mothypotesen ges nu av  $H_1 : \mu < 5$ . Vi kan använda samma teststorhet och om  $H_0$  är sann så är

$$Z = \frac{\bar{X} - 5.0}{0.1/\sqrt{9}} \sim N(0, 1).$$

Vi söker en gräns  $a$  så att  $\bar{X} < a$  med sannolikheten  $\alpha = 0.1336$  om  $H_0$  är sann.



Det är tydligt att

$$\bar{X} < a \Leftrightarrow Z = \frac{\bar{X} - 5.0}{0.1/3} < \frac{a - 5.0}{0.1/3},$$

så

$$\begin{aligned} 0.1336 &= P\left(Z < \frac{a - 5.0}{0.1/3}\right) \Leftrightarrow \frac{a - 5.0}{0.1/3} = \Phi^{-1}(0.1336) = -1.1095 \\ &\Leftrightarrow a - 5.0 = -0.0370. \end{aligned}$$

Det sökta värdet blir alltså  $a = 4.9630$ . Detta test blir alltså mer känsligt för att materialet är för tunnt än det föregående. Om den nya beställaren har samma tolerans för fel som de tidigare är det kanske mer strategiskt att istället sänka signifikansnivån till hälften.

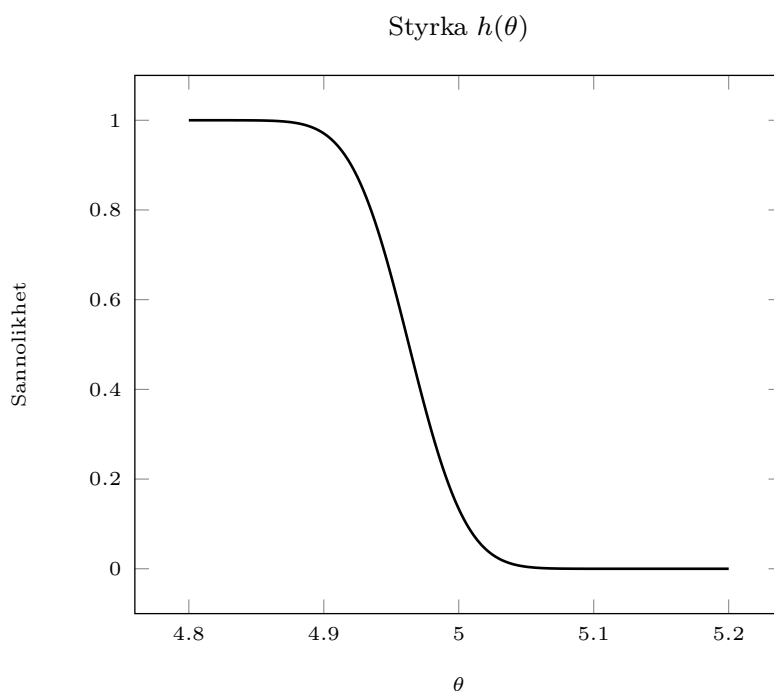


### Exempel

Föregående hypotestest har en lite udda signifikansnivå. Hur ser styrkefunktionen ut?

**Lösning.** Styrkefunktionen definieras enligt

$$\begin{aligned} h(\theta) &= P(H_0 \text{ förkastas} \mid \mu = \theta) = P\left(\bar{X} < 4.9630 \mid \bar{X} \sim N\left(\theta, \frac{0.1^2}{9}\right)\right) \\ &= \Phi\left(\frac{4.9630 - \theta}{0.1/3}\right) = \Phi(148.89 - 30\theta). \end{aligned}$$



### Exempel

En nyanställd i fabriken (med en kurs i statistisk inferens i bagaget) påtalar att det kanske är olämpligt att anta att variansen är känd och att man borde skatta den från mätningen. Vid en mätning fick man stickprovsvariansen 0.0144, vad ger testet  $|\bar{x} - 5.0| > 0.05$  för signifikansnivå i denna situation?

**Lösning.** Vi testar således  $H_0 : \mu = 5.0$  mot  $H_1 : \mu \neq 5.0$  och som testvariabel blir

$$T = \frac{\bar{X} - 5.0}{S/\sqrt{9}} \sim t(8)$$

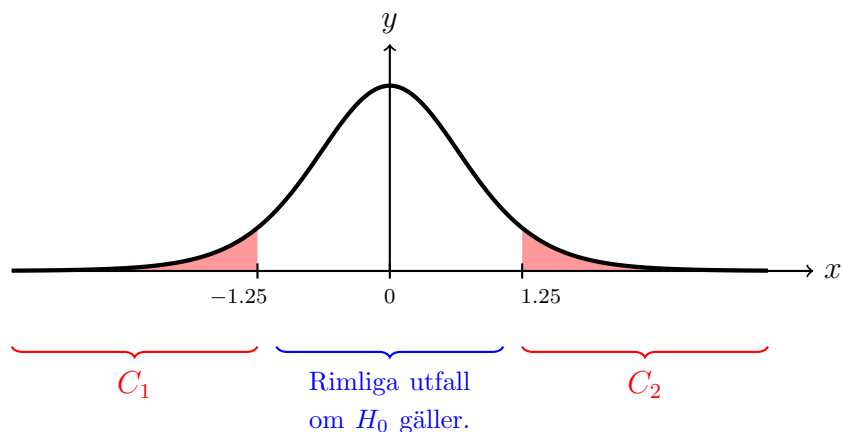
om  $H_0$  är sann. Analogt med första exemplet måste då

$$\left| \frac{s}{3} t \right| > 0.05 \quad \Leftrightarrow \quad |t| > \frac{0.15}{s},$$

vilket ger det kritiska området

$$C = \left\{ t \in \mathbf{R} : |t| > \frac{0.15}{s} \right\} = \{t \in \mathbf{R} : |t| > 1.25\}$$

i vårt fall. Fördelningen för  $T$  är symmetrisk lik normalfördelningen, så situationen är snarlik.



Så  $p$ -värdet kan om fördelningen ser symmetrisk ut enligt ovan beräknas enligt

$$\begin{aligned} p &= P(T \leq -1.25) + P(T \geq 1.25) = 2P(T \leq -1.25) \\ &= 2F_T(-1.25) = 2 \cdot 0.1233 = 0.2466. \end{aligned}$$

Här använde vi `tcdf(-1.25,8)` i MATLAB (vi har inga tabeller i formelsamlingen för att slå på  $t$ -fördelningar i ”den riktningen”).

Den bästa signifikansnivån vi kan välja är alltså i princip  $\alpha = 0.25$ . Mindre lyckat! Testet kanske behöver ändras.