Föreläsning 5: Hypotesprövningar

Johan Thim (johan.thim@liu.se)

12 mars 2020

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

Teststorhet/Testvariabel

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

För att avgöra om vi ska förkasta H_0 väljer vi en **signifikansnivå** α 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.

Kristiskt område, signifikansnivå

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

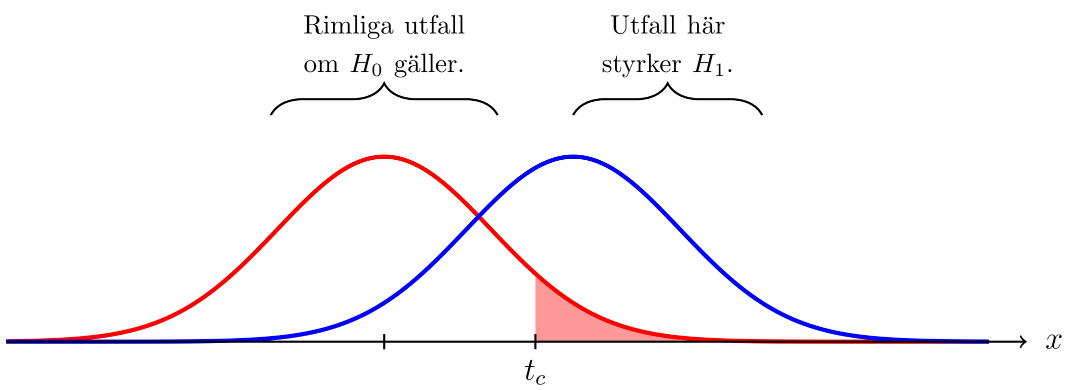
$$t(x_1, \ldots, x_n) \in C.$$

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

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

kallas för testets **signifikansnivå**.

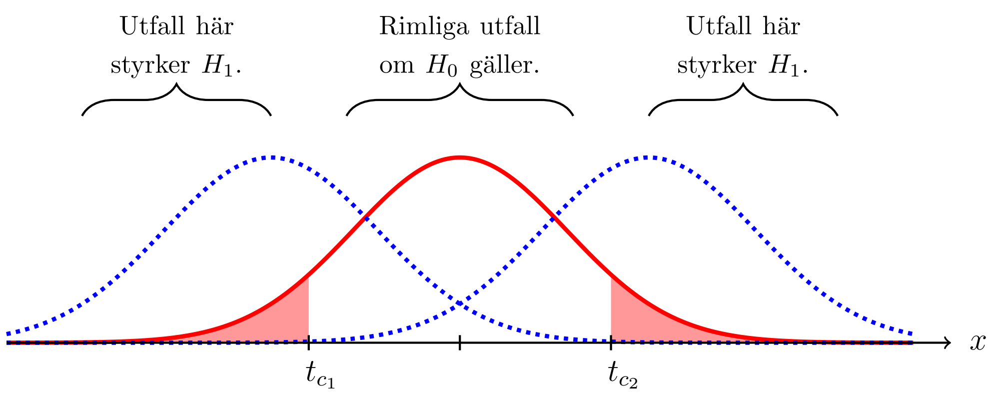
Det kritiska området består alltså av värden som är för extrema för att vara troliga under förutsättningen att nollhypotesen gäller.
Låt oss ställa upp ett hypotestest för väntevärdet för fördelningen F enligt $H_0 : \mu = \mu_0$ mot $H_1 : \mu > \mu_0$. Vi vill således styrka att det verkliga väntevärdet är större än μ_0.



Den röda kurvan är täthetsfunktionen för $t(X_1, \ldots, X_n)$ 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 μ_0 (inte nödvändigtvis att det verkliga väntevärdet är större).



De blå kurvorna är potentiella verkliga fördelningar för $t(X_1, \ldots, X_n)$ 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 $t(X_1, \ldots, X_n)$ 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 beter man sig bedrägligt.

1.2 Styrka, fel och p-värde

Så säg att vi har valt en 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 θ_0). Detta blir dock svårt att uppfylla, men hur ser sannolikheten ut för att korrekt förkasta H_0 om vi låter θ variera? Detta brukar kallas för testets **styrka**.

Styrka

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

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

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

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 $h(\theta_0) = \alpha$.
Uppenbarligen finns det en risk att vi tar fel beslut. Denna riska kan delas upp i två olika typer.

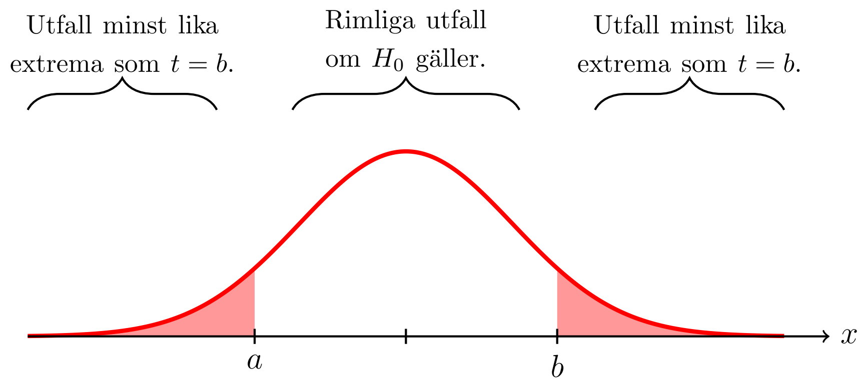
Fel av typ I och II

Definition. Att förkasta H_0 då H_0 är sann kallas **fel av typ I** och har sannolikheten α. 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**.

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 $t(x_1, \ldots, x_n) = b$ 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

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

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

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

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%? Vad är styrkan om sannolikheten för krona är 3/10?

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

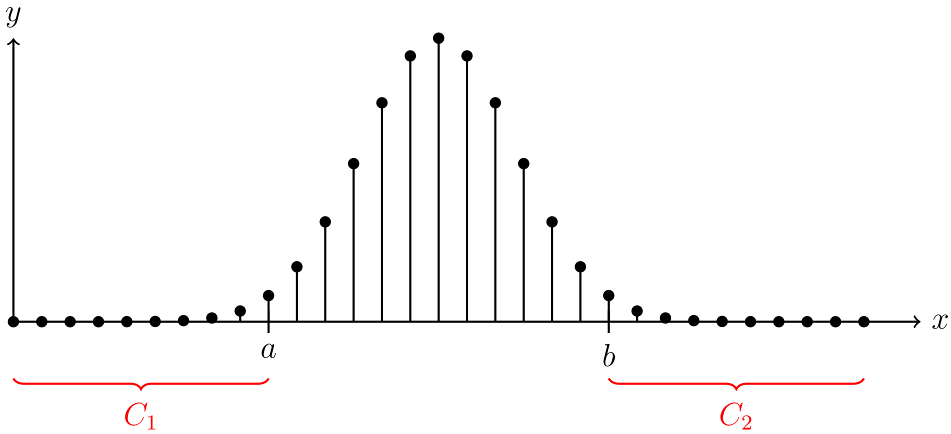
$$H_0 : p = \frac{1}{2}$$

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

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

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.

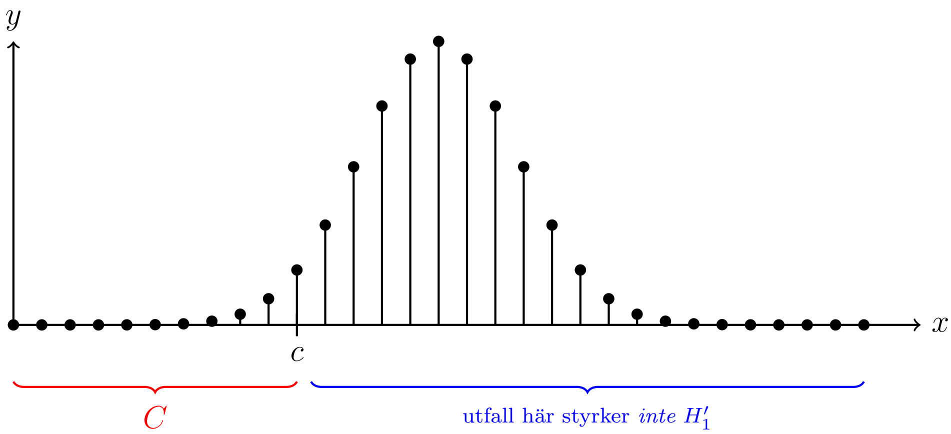
Eftersom $x = 10 \notin C$ kan vi inte dra någon slutsats, utan myntet kan mycket väl vara ärligt. Vi kan således *inte* förkasta H_0 (vilket inte på något sätt betyder att H_0 är sann).
Styrkan vid $p = 0.3$ blir

$$\begin{aligned} h(0.3) &= P(H_0 \text{ förkastas} \,|\, p = 0.3) = P(X \in C \,|\, p = 0.3) \\ &= \sum_{x=0}^{9} \binom{30}{x} 0.3^x 0.7^{30-x} + \sum_{x=21}^{30} \binom{30}{x} 0.3^x 0.7^{30-x} = 0.5888 + 7.28 \cdot 10^{-6} = 0.5888. \end{aligned}$$

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

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

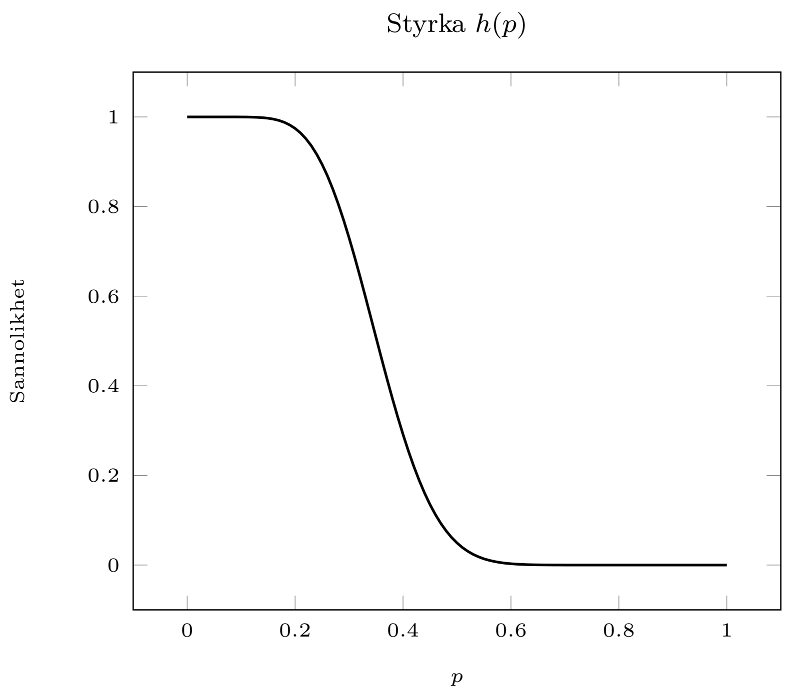
så ser vi att $c = 10$ är nödvändigt. Därmed blir

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

och vår observation $x = 10 \in C$. Alltså kan vi förkasta H_0 och anse att H_1' är styrkt.
Styrkan vid $p = 0.3$ blir

$$\begin{aligned} h(0.3) &= P(H_0 \text{ förkastas} \,|\, p = 0.3) = P(X \in C \,|\, p = 0.3) \\ &= \sum_{x=0}^{10} \binom{30}{x} 0.3^x 0.7^{30-x} = 0.7304. \end{aligned}$$

Notera alltså att styrkan *beror* på mothypotesen! Ganska naturligt när man tänker efter, men det är lätt att tro att styrkan för ett test bara har med nollhypotesen att göra. Det är alltså helt fel. Vi kan även låta MATLAB räkna ut styrkefunktionen för alla $p \in [0, 1]$ för att se hur det ser ut.



3 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.

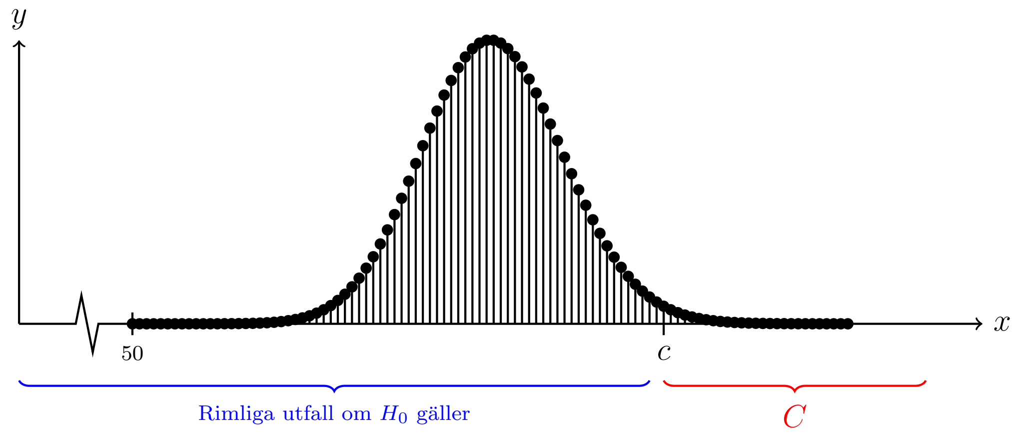
Exempel

Antalet datapaket till en server kan betraktas som en Poissonprocess $X(t)$ med en okänd intensitet λ. 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 $\mu = 50 \cdot 2 = 100$ (tusen) 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 $p(k)$, $k = 0, 1, 2, \ldots$, 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

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

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 $X \overset{\text{appr.}}{\sim} N(\mu, \mu)$ (variansen är μ). 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

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

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

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

4 Normalapproximation – Generellt

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

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

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 $Z \overset{\text{appr.}}{\sim} N(0,1)$ 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.

5 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 $p = p_1 = p_2$

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

Faktum är att detta är ML-skattningen (om H_0 är sann) och därmed har den bra egenskaper såsom konsistens. Vad gäller fördelningen för $\widehat{P}$ 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

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

och

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

då $n_1 + n_2 \to \infty$, så skattningen av p är väntevärdesriktig och konsistent. För att testa H_0 använder vi $\widehat{P_1} - \widehat{P_2}$, och om H_0 är sann så gäller att

$$\widehat{P_1} - \widehat{P_2} \overset{\text{appr.}}{\sim} N\left(0,\, \widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

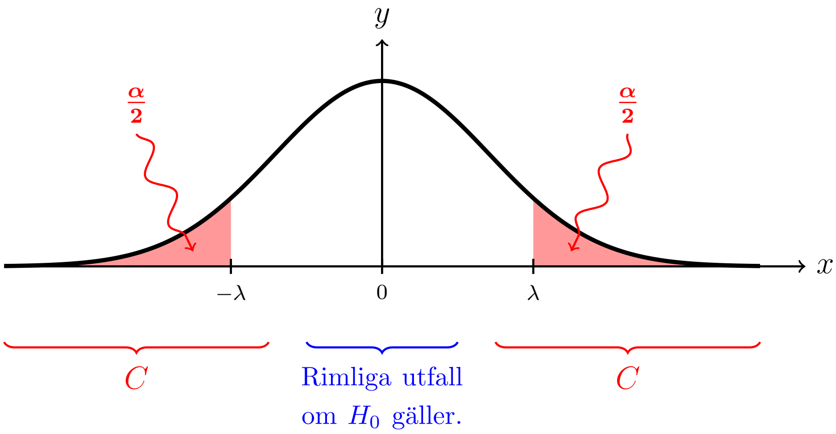
Eftersom vi inte känner p exakt använder vi skattningen $\widehat{p}$ ovan i uttrycket för variansen (eller vi ersätter standardavvikelsen med medelfelet). Vi kan även gå över i standardiserad form så vi känner igen oss:

$$Z = \frac{\widehat{P_1} - \widehat{P_2}}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \overset{\text{appr.}}{\sim} N(0,1).$$

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

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

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

$$Z = \frac{\widehat{P_1} - \widehat{P_2}}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{500} + \frac{1}{400}\right)}} = \frac{\widehat{P_1} - \widehat{P_2}}{0.0293} \overset{\text{appr.}}{\sim} N(0,1).$$

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.

6 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). Laboratorieansvarig 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 $\widehat{P} = \dfrac{X}{1000}$, men enklare är att direkt nyttja X. Om H_0 är sann så gäller att

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

Det kritiska området väljs som

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

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

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

och i tabell (eller med `k = 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

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

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 $\widehat{p} = 10/1000 = 0.01$ hamnar vi precis kring den gränsen så osäkerheten är stor. Använder vi leverantörens $p = 0.005$ blir det betydligt under. Alltså inget att rekomendera. Men om vi envisas så skulle

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

Om vi antar att H_0 är sann skulle då

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

återigen som en tveksam approximation. Kritiskt område ges av

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

så

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

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 $P(X > k) = 0.0053$.