Föreläsning 9: Hypotesprövning

Matematisk statistik

David Bolin
Chalmers University of Technology
Maj 7, 2015

Konfidensintervall

Konfidensintervall

Låt $X_1, \ldots, X_n$ vara slumpvariabler med en fördelning som har θ som en parameter med θ_0 som sant okänt värde. Ett $100(1-\alpha)\%$ konfidensintervall för θ med konfidensgraden $1-\alpha$ är ett intervall $[a(X_1, \ldots, X_n), b(X_1, \ldots, X_n)]$ sådant att

$$\mathsf{P}(a \leq \theta_0 \leq b) = 1 - \alpha.$$

- (a, b) är ett slumpmässigt intervall, eftersom a och b är slumpvariabler som beror av $X_1, \ldots, X_n$.
- Konfidensintervallet ska alltså tolkas som att om vi gör upprepade mätningar av $X_1, \ldots, X_n$ och bildar konfidensintervallet för alla dessa mätningar så kommer $100(1-\alpha)\%$ av dessa intervall täcka det sanna värdet θ_0.

Konfidensintervall för μ i normalfördelningen

Känd varians

Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(\mu, \sigma^2)$ där σ^2 är känd. Intervallet

$$I_\mu = (\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$$

är ett konfidensintervall för μ med konfidensgrad $1 - \alpha$.

Okänd varians

Låt $X_1, \ldots, X_n$ vara oberoende $\mathrm{N}(\mu, \sigma^2)$ där σ^2 är okänd. Intervallet

$$I_\mu = (\bar{X} - t_{\alpha/2}(n-1)S/\sqrt{n}, \bar{X} + t_{\alpha/2}(n-1)S/\sqrt{n})$$

är ett konfidensintervall för μ med konfidensgrad $1 - \alpha$. Här är $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ och $t_{\alpha/2}(n-1)$ är $\alpha/2$-kvantilen i t-fördelningen med $n - 1$ frihetsgrader.

Centrala gränsvärdessatsen (CGS)

Kom ihåg att CGS säger att $\bar{X}$ är approximativt $\mathrm{N}(\mu, \sigma^2/n)$-fördelad om n är stort.

- Om vi har ett stickprov med känd varians σ^2 kommer

$$I_\mu = (\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$$

  vara ett approximativt konfidensintervall för väntevärdet μ.
- Om vi inte känner variansen kan vi skatta σ med S, men det är då viktigt att n är stort och att fördelningen för X_i inte är allt för tungsvansad om skattningen ska bli någorlunda bra.
- Eftersom n är stor kommer i detta fall $t_{\alpha/2}(n-1) \approx z_{\alpha/2}$, så vi använder

$$I_\mu = (\bar{X} - z_{\alpha/2}S/\sqrt{n}, \bar{X} + z_{\alpha/2}S/\sqrt{n})$$

  som ett approximativt konfidensintervall i fallet då σ är okänd.

$\chi^2(n)$-fördelningen

$\chi^2(n)$-fördelningen

En slumpvariabel är $\chi^2(n)$-fördelad (chitvåfördelad med n frihetsgrader) om den har täthetsfunktion

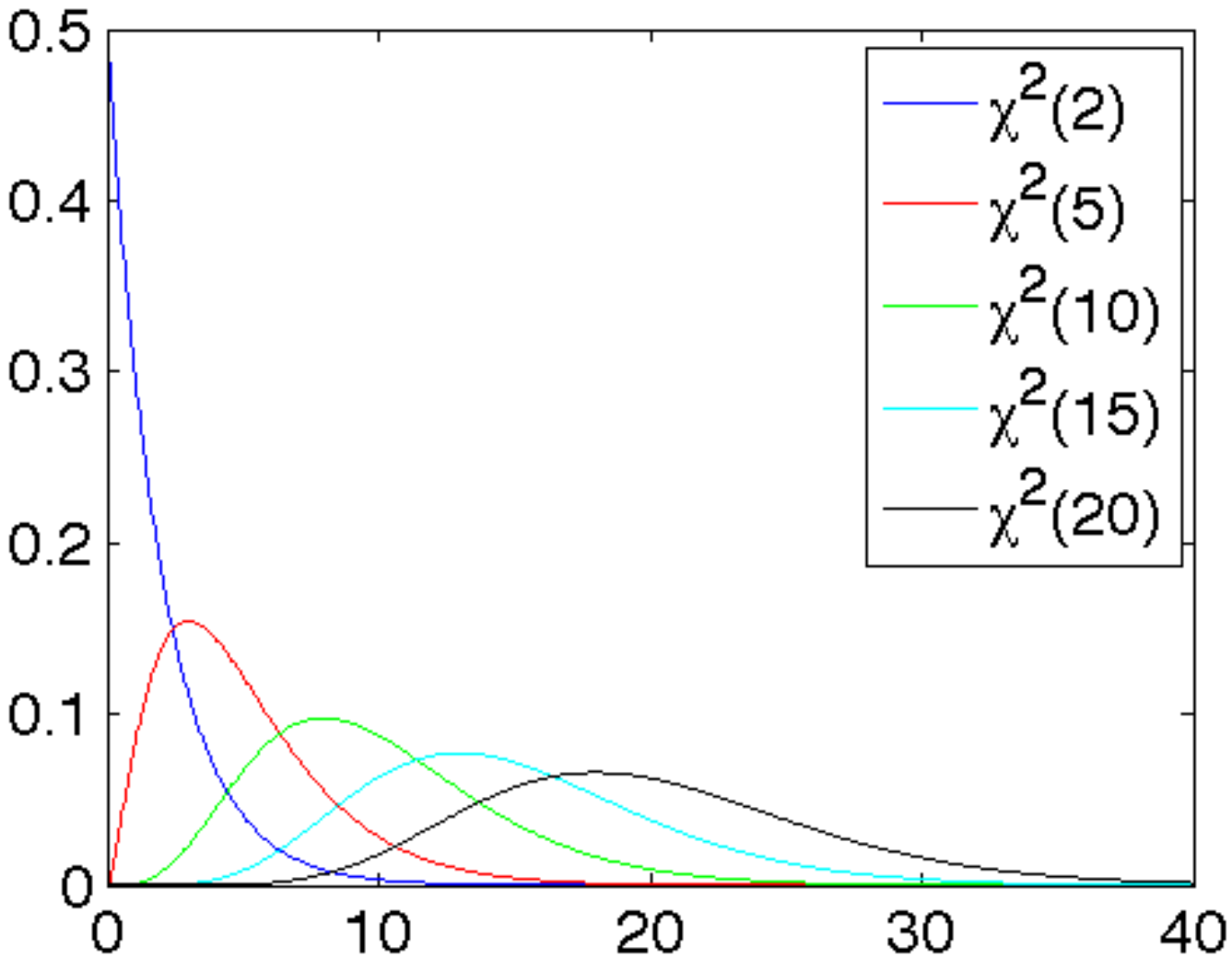
$$f(x) = \begin{cases} \frac{x^{n/2-1}e^{-x/2}}{2^{k/2}\Gamma(\frac{n}{2})}, & \text{om } x \geq 0, \\ 0 & \text{om } x < 0. \end{cases}$$

Om Z_i är oberoende $\mathrm{N}(0,1)$, så gäller att

$$\sum_{i=1}^{n} Z_i^2$$

är $\chi^2(n)$-fördelad.

$\chi^2(n)$-fördelningen



Konfidensintervall för σ^2 i normalfördelningen

Konfidensintervall för σ^2

Om $X_1, \ldots, X_n$ är oberoende $\mathrm{N}(\mu, \sigma^2)$ fås ett konfidensintervall med konfidensgrad $1 - \alpha$ för σ^2 som

$$I_{\sigma^2} = \left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)} \right).$$

Här är $\chi^2_{\alpha/2}(n-1)$ $\alpha/2$-kvantilen i χ^2 fördelningen med $n-1$ frihetsgrader.

Konfidensintervall för σ^2 i normalfördelningen

Konfidensintervall för σ

Om $X_1, \ldots, X_n$ är oberoende $\mathrm{N}(\mu, \sigma^2)$ fås ett konfidensintervall med konfidensgrad $1 - \alpha$ för σ som

$$I_\sigma = \left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}} \right).$$

Här är $\chi^2_{\alpha/2}(n-1)$ $\alpha/2$-kvantilen i χ^2 fördelningen med $n-1$ frihetsgrader.

En viktig anmärkning här är att till skillnad från konfidensintervall för väntevärdet så är konfidensintervall för variansen mycket känsliga för avvikelser från normalfördelningen.

Ensidiga konfidensintervall

Ibland är man endast intresserad av en trolig övre eller undre gräns för parametern. Vi kan då bilda ett ensidigt konfidensintervall. Låt $X_1, \ldots, X_n$ vara slumpvariabler med en fördelning som har θ som en parameter med θ_0 som sant okänt värde.

- Ett undre begränsat ensidigt $100(1-\alpha)\%$ konfidensintervall för θ med konfidensgraden $1-\alpha$ är ett intervall $[a(X_1, \ldots, X_n), \infty]$ sådant att

$$\mathsf{P}(a \leq \theta_0) = 1 - \alpha.$$

- Ett övre begränsat konfidensintervall ges på samma sätt av ett intervall $[-\infty, b(X_1, \ldots, X_n)]$ sådant att

$$\mathsf{P}(\theta_0 \leq b) = 1 - \alpha.$$

I de fallen vi tittat på få ensidiga konfidensintervall genom att endast beräkna en av gränserna samt att vi byter $\alpha/2$-kvantilen till en α-kvantil.

Hypotesprövning

Ett viktigt problem inom statistiken är att kunna testa om en teori eller en hypotes är sann. Exempel på sådanna frågeställningar kan vara

- Ger ett nytt läkemedel någon effekt?
- Dör rökare tidigare än icke rökare?
- Har mätinstrumentet ett systematiskt fel?

Svaret som den statistiska analysen ger kan vara

1. att hypotesen styrks eller bekräftas av försöket, och man får då också ett mått på hur säker denna slutsats är.
2. att datan inte ger något stöd för hypotesen.

Hypotesprövning

Vi vill testa *nollhypotesen*

$$H_0 : \theta = \theta_0$$

mot *mothypotesen*

$$H_1 : \theta \neq \theta_0.$$

Om testet bekräftar mothypotesen säger vi att H_0 förkastas till förmån för H_1. Man brukar då också säga att θ är signifikant skilt från θ_0.

Principen för testet bör vara att vi beräknar en skattning θ^* av θ och förkastar H_0 om θ^* är långt från θ_0. Vi ställs då inför frågorna

1. Vad innebär att θ^* är långt från θ_0?
2. Hur kan vi ange säkerheten i slutsatsen vi drar?

Konfidensintervallmetoden

Om vi bildar ett 95% konfidensintervall I_θ för θ kan vi direkt använda detta för att göra hypotestestet.

- Eftersom intervallet är beräknat så att det ska täcka det sanna värdet på θ i 95% av fallen så kan vi förkasta H_0 om intervallet inte innehåller θ_0.
- Om intervallet däremot innehåller θ_0 kan vi inte förkasta H_0, dvs påstå att $\theta \neq \theta_0$.

Konfidensgraden vi använder när vi beräknar konfidensintervallet är också konfidensgraden för hypotestestet vi gör.

- Det har blivit standard inom många områden att använda 95% konfidensgrad, men vi kan också använda en högre eller en lägre konfidensgrad.

Konfidensgrad

Ett konfidensintervall med konfidensgrad $100(1-\alpha)\%$ är gjort så att det täcker det sanna värdet av θ med sannolikhet $1-\alpha$.

Om H_0 är sann, dvs $\theta = \theta_0$, finns det alltså en risk α att intervallet inte täcker θ_0 och vi drar då felaktigt slutsatsen att $\theta \neq \theta_0$.

Denna risk är det vi kallar för *felrisk* eller *signifikansnivå*.

Definition

Felrisken eller signifikansnivån definieras som

$$\alpha = \mathrm{P}(H_0 \text{ förkastas } | H_0 \text{ sann}).$$

Översatt till konfidensintervall är signifikansnivån

$$\alpha = \mathrm{P}(\theta \notin I_\theta \text{ om } \theta = \theta_0) = \mathrm{P}(\theta_0 \notin I_\theta).$$

Alltså fås signifikansnivån som 1 - konfidensgraden.

Typ 2 fel

Felet vi gör om vi förkastar H_0 trots att den är sann brukar kallas för ett "Typ 1 fel". Den andra sortens fel vi kan göra är "Typ 2 fel":

Definition

Under hypotesprövning säger vi att vi gör ett typ 2 fel om vi inte förkastar H_0 trots att H_1 är sann. Sannolikheten för att göra ett typ 2 fel brukar betecknas med β:

$$\beta = \mathrm{P}(H_0 \text{ förkastas inte } | H_1 \text{ sann}).$$

Vi sätter själva sannolikheten för att göra ett typ 1 fel när vi designar testet. Typ 2 fel är lite svårare att hantera eftersom sannolikheten att göra ett typ 2 fel beror på mothypotesen. Vi återkommer till detta senare.

De olika utfallen av ett hypotestest

För att sammanfatta kan fyra olika saker hända när vi gör ett hypotestest:

1. H_0 är sann och hypotestestet förkastar inte H_0.
2. H_0 är sann och hypotestestet förkastar H_0. Detta är ett typ 1 fel och har enligt konstruktionen av testet sannolikhet α att inträffa.
3. H_1 är sann och hypotestestet förkastar H_0.
4. H_1 är sann och hypotestestet förkastar inte H_0. Detta är ett typ 2 fel och sannolikheten för att detta inträffar brukar betecknas med β.

En viktig observation

I fallet då vi inte kan förkasta H_0 bör det understrykas att testet *inte* visar att $\theta = \theta_0$, eftersom både θ_0 och många andra värden täcks av konfidensintervallet.

Frågan om H_0 eller H_1 är sann lämnas alltså då öppen. Det finns alltså en asymmetri mellan H_0 och H_1 eftersom vi endast kan använda testet för att visa att den ena hypotesen är sann.

Av tradition väljer man alltid den hypotes man vill kunna visa är sann som H_1.

Hypotesprövning med teststorhet

Om vi kan förkasta H_0 med signifikansnivå 0.05 kan det vara intressant att se om vi också kan förkasta med en lägre signifikansnivå, tex 0.01.

En nackdel med konfidensintervallsmetoden är att vi då måste beräkna om konfidensintervallet för varje signifikansnivå vi vill undersöka.

Ett alternativ är att göra testet baserat på en teststorhet, vilket är den generella metoden för hypotestest. Denna metod har två fördelar:

1. Beräkningarna blir ofta enklare.
2. Man får ett exakt mått på säkerheten i slutsatsen, ett så kallat p-värde.

Teststorheter

Definition

En teststorhet $T = T(X_1, \ldots, X_n)$ är en funktion av observationerna, och alltså en slumpvariabel. $T_{obs} = T(x_1, \ldots, x_n)$ är ett observerat värde av teststorheten för givna observationer.

I allmänhet innehåller teststorheten en skattning θ^* av parametern θ. Till exempel för test av väntevärdet i en normalfördelning med okänd varians kan vi använda

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Om vi känner variansen kan vi istället använda

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Dessa teststorheter känns igen som de slumpvariabler vi använde för att beräkna konfidensintervall för normalfördelningen.

p-värden

Givet att vi har formulerat vår teststorhet gör vi testet genom att beräkna ett p-värde.

Definition

p-värdet eller signifikanssannolikheten definieras som sannolikheten under nollhypotesen att vi får ett värde $|T|$ som är lika stort eller större än det observerade värdet $|T_{obs}|$.

Vi vill använda en storhet T som vi känner fördelningen av under H_0, så att vi kan beräkna p-värdet.

Om fördelningen för T är symmetrisk ges i allmänhet p-värdet av

$$p = 2(1 - F_T(|T_{obs}|))$$

där F_T är fördelningsfunktionen för teststorheten.

p-värden (forts.)

I fallet med normalfördelning med känd varians har vi att T under H_0 är $\mathsf{N}(0,1)$-fördelad, så vi har att

$$p = \mathsf{P}(|T| \geq |T_{obs}|) = 2 \cdot \mathsf{P}(T \geq |T_{obs}|) = 2(1 - \Phi(|T_{obs}|)).$$

Vi förkastar H_0 om p är litet, och det exakta värdet på p är den minsta signifikansnivå vi kan ha för att förkasta H_0.

Om vi vill göra ett test med en fix signifikansnivå α, tex 0.05, behöver vi alltså endast beräkna p och sedan förkasta H_0 om p$< \alpha$.

Kritiskt område

Om vi endast är intresserade av att göra testet med en viss signifikansnivå α kan vi också beräkna det så kallade kritiska området för teststorheten istället för att beräkna p-värdet.

Definition

Givet en signifikansnivå α definierar vi det kritiska området C_α som de värden på teststorheten T som leder till att man förkastar H_0 på nivån α.

Vi förkastar H_0 på nivån α om $p < \alpha$, vilket ekvivalent kan uttryckas som $|T| > T_{\alpha/2,H_0}$, där $T_{\alpha/2,H_0}$ är $\alpha/2$-kvantilen i T:s fördelning under H_0. Därför har vi $C_\alpha = \{T : |T| > T_{\alpha/2,H_0}\}$.

Kritiskt område för normalfördelningen

Med test för väntevärdet hos normalfördelade data får vi:

- Om variansen är känd: T under H_0 är $\mathsf{N}(0,1)$-fördelad, förkasta H_0 på nivån α om $|T| > z_{\alpha/2}$.
- Om variansen är okänd: T under H_0 är $t(f)$-fördelad, förkasta H_0 på nivån α om $|T| > t_{\alpha/2}(f)$.

Ensidiga test

I vissa situationer kan vi vilja testa om det observerade värdet ligger för högt respektive för lågt i förhållande till referensvärdet.

Vi vill då göra ett så kallat ensidigt test.

Vid ett ensidigt test förkastar vi H_0 endast om vi kan hävda att parametern är för stor (alternativt för liten). Nollhypotesen är som tidigare

$$H_0 : \theta = \theta_0$$

men mothypotesen är nu antingen

$$H_1 : \theta > \theta_0$$

eller

$$H_1 : \theta < \theta_0$$

Ensidiga test (forts.)

Om vi använder ett konfidensintervall för att göra testet ska vi nu använda ett ensidigt konfidensintervall

- om $H_1 : \theta > \theta_0$: $I_\theta = \{a, \infty\}$.
- om $H_1 : \theta < \theta_0$: $I_\theta = \{-\infty, b\}$.

Vi förkastar H_0 om intervallet inte täcker θ_0.

Alternativt använder vi oss av en teststorhet och vi beräknar då p-värdet som

- om $H_1 : \theta > \theta_0$: $p = \mathsf{P}_{H_0}(T \geq T_{obs})$.
- om $H_1 : \theta < \theta_0$: $p = \mathsf{P}_{H_0}(T \leq T_{obs})$.

och förkastar sedan H_0 om $p \leq \alpha$.

Kritiskt område

Använder vi oss av ett test med en fix signifikansnivå kan vi också beräkna det kritiska området.

Med test för väntevärdet hos normalfördelade data får vi nu:

- om $H_1 : \theta > \theta_0$:
  - Om σ är känd: Förkasta H_0 på nivån α om $T > z_\alpha$.
  - Om σ är okänd: Förkasta H_0 på nivån α om $T > t_\alpha(f)$.
- om $H_1 : \theta < \theta_0$:
  - Om σ är känd: Förkasta H_0 på nivån α om $T < -z_\alpha$.
  - Om σ är okänd: Förkasta H_0 på nivån α om $T < -t_\alpha(f)$.

Observera att vi har bytt $\alpha/2$ mot α här, precis som vi gör när vi beräknar ensidiga konfidensintervall.

Styrka

Kom ihåg att felet vi gör då vi vid hypotesprövning inte förkastar nollhypotesen i fallet då mothypotesen faktiskt är sann kallas för ett ”Typ 2 fel”.

Sannolikheten för att göra ett sådant fel betecknas med β. Sannolikheten att *inte* göra ett typ 2 fel kallas för testets *styrka*, och ges alltså av $1 - \beta$.

Antag att vi vill testa om det förekommer en systematisk avvikelse från θ_0. Om den verkliga avvikelsen är liten är det inte så viktigt om testet ger ett icke-signifikant resultat. Om den verkliga avvikelsen däremot är så stor att den är av praktisk betydelse så vill vi att testet ska ge utslag om detta.

Vi bryr oss alltså inte så mycket om typ 2 fel om den systematiska avvikelsen är liten, men vi vill inte göra ett typ 2 fel om avvikelsen är stor.

Styrkefunktion

Typ 2 fel är lite svårare att hantera eftersom de beror på mothypotesen men ett sätt att visualisera dessa är att beräkna testets *styrkefunktion*

Definition

Vi vill testa $H_0 : \theta = \theta_0$ och har bestämt en teststorhet T och en signifikansnivå α. Testets styrkefunktion $\pi(\theta)$ definieras som

$$\pi(\theta) = \mathrm{P}(T \in C_\alpha | \theta) = \mathrm{P}(H_0 \text{ förkastas} | \text{ parametervärdet är } \theta)$$

Testets styrkefunktion ger alltså sannolikheten för att testet ger ett signifikant resultat om det sanna värdet är θ, vilket alltså är testets styrka vid värdet θ.

Sats

Antag att vi har n observationer av en normalfördelning $\mathrm{N}(\mu, \sigma^2)$ med känd varians och att vi vill testa $H_0 : \mu = \mu_0$. Testets styrka ges då av

- Om $H_1 : \mu > \mu_0$:

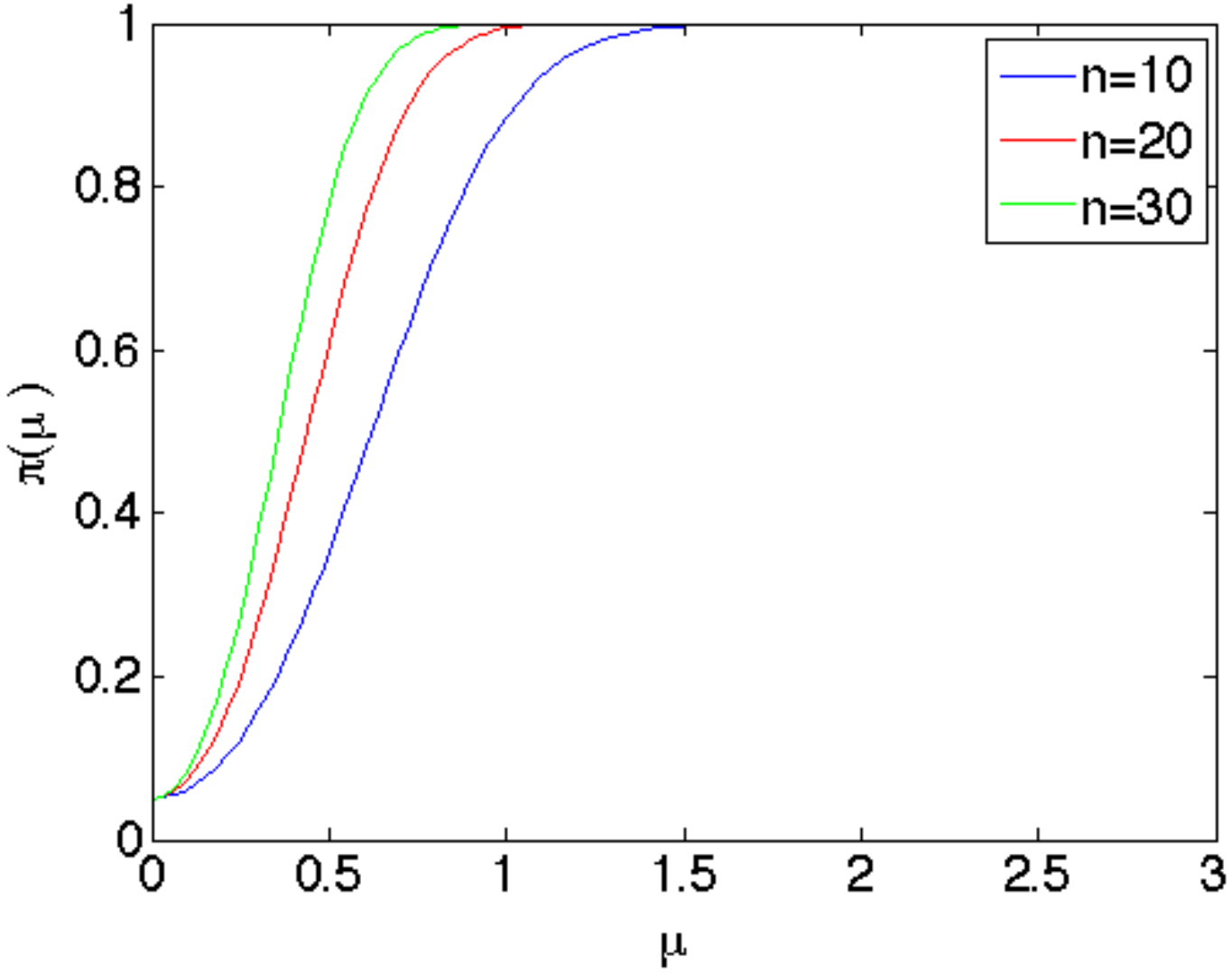
$$\pi(\mu) = 1 - \Phi\left(z_\alpha - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right)$$

- Om $H_1 : \mu < \mu_0$:

$$\pi(\mu) = \Phi\left(-z_\alpha - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right)$$

- Om $H_1 : \mu \neq \mu_0$:

$$\pi(\mu) = 1 + \Phi\left(-z_{\alpha/2} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) - \Phi\left(z_{\alpha/2} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right)$$



Styrkefunktion för test av $H_0 : \mu = 0$ mot $H_1 : \mu \neq 0$ för N$(\mu, 1)$-fördelningen på nivå $\alpha = 0.05$.

Hur många observationer krävs?

Det vi kan se från styrkefunktionerna i de olika fallen är att testet har lättare att upptäcka en avvikelse $\mu - \mu_0$ om

- Risken för typ 1 fel, α, är stor.
- Antalet observationer, n, är stort.
- Variansen σ^2 är liten.

Om vi bestämmer oss för en viss avvikelse $\mu - \mu_0$ vi vill att testet ska upptäcka samt fixerar signifikansnivån α kan vi använda styrkefunktionen för att beräkna hur många observationer vi bör göra för att få en given styrka $1 - \beta$ på testet.

Sats

Antag att vi har normalfördelade observationer $\mathrm{N}(\mu, \sigma^2)$ och önskar testa $H_0 : \mu = \mu_0$. Om vi vill att testet ska ge utslag för avvikelsen $\mu - \mu_0$ med sannolikhet $1 - \beta$ ska antal observationer väljas som:

- Känd varians:

$$\text{Ensidigt test: } n = \frac{\sigma^2}{(\mu - \mu_0)^2}(z_\alpha + z_\beta)^2$$

$$\text{Tvåsidigt test: } n = \frac{\sigma^2}{(\mu - \mu_0)^2}(z_{\alpha/2} + z_\beta)^2$$

- Om variansen skattas med s^2:

$$\text{Ensidigt test: } n \approx \frac{\sigma^2}{(\mu - \mu_0)^2}(z_\alpha + z_\beta)^2 + z_\alpha^2/2$$

$$\text{Tvåsidigt test: } n \approx \frac{\sigma^2}{(\mu - \mu_0)^2}(z_{\alpha/2} + z_\beta)^2 + z_{\alpha/2}^2/2$$