SF1914/SF1916: Sannolikhetsteori och statistik
Föreläsning 13
Hypotesprövning.

Tatjana Pavlenko

5 oktober 2018

KTH Matematik

Plan för dagens föreläsning

- Begrepp inom hypotesprövning (rep.)
- Tre metoder för att avgöra om H_0 ska förkastas.
- Styrkefunktionen (Kap. 13.4).
- Tillämpning på normalfördelning (Kap. 13.6).
- Användning av normalapproximation (Kap. 13.7).

- Vi betraktar den allmänna situationen: vi har ett stickprov $x = (x_1, \ldots, x_n)$ som är utfall av s. v. $(X_1, \ldots, X_n)$ från någon fördelning $F_X(x;\theta)$.

- Vi vill testa en grundhypotes, eller *nollhypotes*, om θ: $H_0 : \theta = \theta_0$ mot en *alternativ hypotes*, H_1, som kan vara *enkel*, $H_1 : \theta = \theta_1$, eller *sammansatt*, t ex $H_1 : \theta > \theta_0$. En hypotes av typ $H_1 : \theta > \theta_0$ eller $H_1 : \theta < \theta_0$ är *ensidig*, medan $H_1 : \theta \neq \theta_0$ är *tvåsidig*.

- För att testa H_0 definierar vi först en *test variabel* eller *teststorhet* , $t_{obs} = t(x)$ som är en observation av motsvarande stickprovsvariabel $t(X)$. Vi kommer också att behöva testets signifikansnivån eller felrisk.

- Def: *Med signifikansnivån*, eller *felrisken*, α för ett test menas

$$\alpha = P(H_0 \,\text{förkastas}) \quad \text{om } H_0 \text{ är sann.}$$

- Testet utformas på så sätt att felrisken blir liten. Vanliga nivåer är $\alpha = 0.05$, $\alpha = 0.01$, $\alpha = 0.001$. Jämför med felrisker vid konfidensintervall konstruktion!

Tre metoder för att avgöra om H_0 ska förkastas.

Testvariabelmetoden (rep.)

- Hitta en teststorhet $t(X)$ och till den *ett kritiskt område* C.
- Testet blir då:

$$\text{Om } \begin{cases} t_{obs.} \in C, & \text{förkasta } H_0 \\ t_{obs.} \notin C, & \text{ej förkasta } H_0, \end{cases}$$

  där det kritiska området C anpassas så att

$$P(H_0 \text{ förkastas}) = P(t(X) \in C) = \alpha \quad H_0 \text{ är sann}$$

  och α är testets signifikansnivå som väljs på förhand.
- Tolkning:

  Om H_0 förkastas så föreligger det ett signifikant avvikelse från nollhypotesen H_0 på nivån α.

Direktmetoden (rep.)

- Direktmetoden baseras också på testvariabel, med det kritiska området preciseras inte utan man avgör bara vilka värden på testvariabel som tyder på att H_1 är sann, t ex $t(X) \geq C$.
- Det *direkta* med direktmetoden är att vi räknar ut ett storhet som direkt kan jämföras med felrisken! Metoden går till på följande sätt:

1. Antag att H_0 är sann (dvs att det parametervärde den specificerar är det rätta värde).
2. Räkna under H_0 ut

$$p\text{-värdet} = P(\text{teststorhet blir minst lika extremt som observerat}).$$

3. Om p-värde är mindre än α (ofta $\alpha = 0.05, 0.01, 0.001$) så förkastas H_0 med felrisk α.

Direktmetoden.

- Tolkning: Om p-värde är litet (ofta $<$ än $0.05, 0.01$ eller 0.001) så tror vi inte på H_0. p-värde ger alltså ett mått på *hur orimlig H_0 är.*
- Beslutsregel:

  om p-värde $< \alpha$ så förkastas H_0.
- Detta i termer av testvariabelmetoden betyder att $t_{obs.}$ hamnar i det kritiska området C.

Konfidensmetoden.

- Konfidensmetoden går till på följande sätt:
  1. Beräkna ett konfidensintervall I_θ för parametern med samma felrisk som önskas för testet, dvs med konfidensgrad $1-\alpha$.
  2. Förkasta $H_0 : \theta = \theta_0$ om $\theta_0 \notin I_\theta$.
- Typen av konfidensintervall som ska beräknas beror på hur den alternativa hypotesen ser ut:
  - om $H_1 : \theta \neq \theta_0$ ska I_θ vara tvåsidigt,
  - om $H_1 : \theta > \theta_0$ ska I_θ vara ensidigt, *nedåt begränsat*,
  - om $H_1 : \theta < \theta_0$ ska I_θ vara ensidigt, *uppåt begränsat*
- Tolkning: under antagande att H_0 är sann, dvs θ_0 är den sanna parametervärde

$$P(H_0 \text{ förkastas} | H_0 \text{ är sann}) = P(t(X) \in C)$$

$$= P(\theta_0 \notin I_\theta) = 1 - (1-\alpha) = \alpha,$$

dvs testnivå är α.

- Gränsen för rattonykterhet är 2‰.
- *Modell:* antag att ett mättning i är en observation av

$$X_i = \mu + \varepsilon_i,$$

där μ är den sanna halten, $\varepsilon_i \in N(0, \sigma)$ (ober. för $i = 1, \dots, n$) är mätfel där σ antas vara känd, låt $\sigma = 0.04$.

- För att avgöra om en person är skyldig till rattonykterhet kan man använda följande hypoteser:

$$\begin{aligned} H_0 : &\quad \mu = 0.2 \,(\text{oskyldig}) \\ H_1 : &\quad \mu > 0.2 \,(\text{skyldig}). \end{aligned}$$

- Vi väljer felrisken, $\alpha = 0.001$, dvs

$$\alpha = P\big(\text{förkasta } H_0 \text{ om } H_0 \text{ är sann}\big).$$

KTH Matematik

- Att H_0 förkastas innebär att vi finner en person skyldig och om H_0 är sann så är man oskyldig. Felrisk (signifikansnivån) blir alltså

$$P(\text{döma en oskyldig}) = 0.001.$$

Denna händelse inträffar i genomsnitt var tusende gång, dvs vi valde ganska låg felrisk.

- Tillämpning av de tre metoderna för att avgöra om H_0 ska förkastas i ex om rattonykterhet, på tavlan.

- Viktig! De tre metoderna för hypotesprövning är ekvivalenta, dvs ger alltid samma resultat. Detta gäller dock endast *exakta* test, inte alltid approximativa.

Två typ av fel vid hypotesprövning.

Hur bra ett test skiljer H_0 från H_1? Det finns två olika fel (felslutsatser) som kan inträffa vid hypotesprövning. Det brukar kallas

- *fel av första slaget*: att förkasta H_0 trots att H_0 är sann, (*typ I-felet*, den valda felrisken eller signifikansnivån, α)
- *fel av andra slaget*: att inte förkasta H_0 trots att H_1 är sann, (*typ II-felet*).

- Exempel: Test förmåga att förkasta H_0 då den inte är sann, t ex hur stor chans har man att klara sig om man är skyldig med en given promillehalt i exempel om rattonykterhet som togs upp? Vi ska i så fall räkna ut

$$P(\text{dömas med alkoholhalten}\, \mu)$$

dvs, sannolikhet att dömas om μ är den sanna värde.

- För detta brukar man ange testets *styrka* som funktion av parameter μ (allmänt θ).

STYRKEFUNKTION.

- Def. *Styrkefunktionen, $h(\theta)$, för ett test definieras som*

$$h(\theta) = P(H_0 \text{ förkastas}) \text{ om } \theta \text{ är det rätta parametersvärde.}$$

- Ett test är bra om

  $h(\theta)$ är stor för alla $\theta \in H_1$. och
  $h(\theta)$ är liten för alla $\theta \in H_0$.

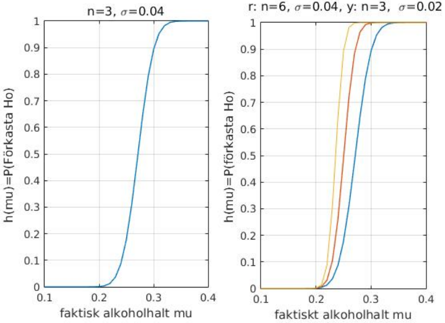
- Med hjälp av styrkefunktionen kan vi räkna ut typ I- och typ II-felet:

  Typ I: $\alpha = P(H_0 \text{ förkastas trots att den är sann}) = h(\theta_0)$
  Typ II: $\beta = P(H_0 \text{ ej förkastas om den ej är sann utan } \theta_1 \text{ är rätt värde}) = 1 - h(\theta_1)$.

- Plottar man styrkefunktionen kan man avläsa de båda felriskerna.
- Exempel: studera testets styrkefunktion för rattonykterhet exempel, på tavlan.

Styrkefunktion (forts.)

För att se hur bra ett test skiljer H_0 från H_1 kan man använda styrkefunktion, $h(\mu)$, som är sannolikhet att H_0 förkastas då μ är rätt värde.



FIGUR: Vänster: styrkefunktionen för alkoholtestet. Höger: samma styrkefuntion samt hur den förändras då man fördubblar antalet mätningar till $n = 6$ eller halverar observationernas $\sigma = 0.02$ (t ex genom att köpa dyrare mätare). I båda fallen ökar testets benägenhet att fälla skyldiga personer.

Tillämpning på normalfördelning: allmänt

Låt $x_1, \ldots, x_n$ vara ett slumpmässigt stickprov från $N(\mu, \sigma)$.

- Vi vill testa
$$H_0 : \mu = \mu_0$$
och använder testvariabel metoden. Då är testvariabel
$$u(x) = \begin{cases} (\bar{x} - \mu_0)/D & \text{om } \sigma \text{ är känd, } D = \sigma/\sqrt{n} \\ (\bar{x} - \mu_0)/d & \text{om } \sigma \text{ är okänd, } d = s/\sqrt{n} \end{cases}$$
- Om H_0 är sann, är dessa kvoter observationer av s.v som är $N(0,1)$ respektive $t(n-1)$ fördelad och vi får kritisk område C med hjälp av motsvarande λ- och t-kvantiler.

KTH Matematik

Beslutsregel erhålls beroende på utseendet hos den alternativa hypotesen H_1:

- (a) Om $H_1 : \mu \neq \mu_0$ (dvs tvåsidig alternativ), så förkastas H_0 om $|u(x)| \geq \lambda_{\alpha/2}$, respektive om $|u(x)| \geq t_{\alpha/2}(n-1)$.
- (b) Om $H_1 : \mu > \mu_0$ (dvs ensidig alternativ), så förkastas H_0 om $u(x) > \lambda_\alpha$, resp. $u(x) > t_\alpha(n-1)$.
- (c) Om $H_1 : \mu < \mu_0$ (dvs ensidig alternativ), så förkastas H_0 om $u(x) < \lambda_\alpha$, resp. $u(x) < t_\alpha(n-1)$.

KTH Matematik

Användning av normalapproximation.

- Har man approximativt normafördelad skattatre, dvs $\theta^* \in AsN(\theta, D(\theta^*))$ då kan man använda testvariabel

$$t(X) = \frac{\theta^* - \theta_0}{D(\theta^*)} \quad \text{om } D(\theta^*) \text{ är känd.}$$

- Om $D(\theta^*)$ innehåller den parameter vi undersöker i H_0 då ska θ_0, dvs värden från H_0 användas, vilket ger $D_0(\theta^*)$ och testvariabeln

$$t(X) = \frac{\theta^* - \theta_0}{D(\theta^*)}.$$

- Annars används

$$t(X) = \frac{\theta^* - \theta_0}{d(\theta^*)}.$$

Kvoten ovan antas ungefär normalfördelad så att man kan utföra ett test på samma sätt som under normalantagande. Viktig! λ-kvantilen används och H_0 förkastas på *approximativt* nivån α.

KTH Matematik