STK1110 Høsten 2026

Kapittel 9 - Uke 38

Hypotesetesting - generelt og i ett utvalg

Oversikt kapittel 9

- Avsnitt 9.1 Innledning - Uke 37
- Avsnitt 9.2 Tester om forventning i ett utvalg - Uke 37-38
- Avsnitt 9.4 P-verdier (alternativ måte å rapportere) (Uke 38)
- Avsnitt 9.3 Tester for binomisk p (Uke 38)
- Avsnitt 9.5 (Bare) Likelihood ratio test (fra s. 547) (Uke 38)
- Avsnitt 9.6 Ytterligere aspekter ved hypotesetesting (Uke 38)

Kort oppsummering av hypotesetesting

for ett utvalg $X_1, \ldots, X_n \sim f(x;\theta)$ uif.

- Interessert i (den ukjente) parameteren θ
- formulert som nullhypotese H_0 vs. alternativ hypotese H_a
- Typiske nullhypoteser $H_0^{'} : \theta \leq \theta_0$, $H_0^{''} : \theta \geq \theta_0$, $H_0^{'''} : \theta = \theta_0$
- og alternative hypoteser $H_a^{'} : \theta > \theta_0$, $H_a^{''} : \theta < \theta_0$, $H_a^{'''} : \theta \neq \theta_0$
- Prøver å konkludere med H_a, men kun med klar evidens mot H_0.
- Velger et nivå $\alpha = \mathrm{P}(\text{ Forkaste } H_0 \text{ gitt at } \theta = \theta_0)$, dvs. gitt at nullhyptesen er riktig
- Finner en testobservator $V = V(X_1, \ldots, X_n)$ for å fastlegge testen
- I praksis kan vi formulere testen ved å forkaste H_0 hvis V er stor nok
- Når testen er fastlagt er vi videre interessert i regne på styrkefunksjonen $\gamma(\theta) = \mathrm{P}(\text{ Forkaste } H_0|\theta)$ for θ gitt ved H_a for å bestemme utvalgsstørrelse n

Eksempel (9.12 i læreboka) Svevestøv fra trafikk

- $n = 9$ observasjoner X_i av d_{50}=medianverdier av svevestøv ved valgte undersøkelsessteder.
- Antar at $X_i \sim \mathrm{N}(\mu, \sigma)$ uif.
- Vil test $H_0 : \mu = \mu_0 = 44$ mot $H_a : \mu \neq 44$
- Velger $\alpha = 0.01$
- Antar først at $\sigma = 20$ er kjent
- Kan benytte testobservator $|Z|$ der $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 44}{20/\sqrt{9}} \sim \mathrm{N}(0, 1)$ under $H_0 : \mu = \mu_0 = 44$
- Da vil vi forkaste med angitt nivå hvis $|Z| > z_{0.005} =$ 99.5-persentil i $\mathrm{N}(0, 1)$, dvs. hvis $|Z| > 2.58$
- I dette tilfellet ble $\bar{x} = \frac{1}{9} \sum_{i=1}^{9} x_i = 68.52$
- og dermed testobservator $|z| = \frac{68.52 - 44}{20/\sqrt{9}} = 3.68 > z_{0.005}$, så H_0 forkastes

Eksempel (9.12 i læreboka) Svevestøv fra trafikk, forts.

- Vi gjorde på forrige slide den urealistiske antagelsen at $\sigma = 20$ var kjent.
- Vi bør erstatte den med empirisk standardavvik S og ta hensyn til dette i testen
- Men vi vet at $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ når $\mathrm{E}(X_i) = \mu_0$
- Med ukjent σ forkastes derfor H_0 hvis $|T| > t_{0.005} =$ 99.5-persentilen i t_{n-1}-fordelingen.
- og med $n = 9$ og $n - 1 = 8$ frihetsgrader finner vi $t_{0.005} = 3.36$
- For disse dataene ble $S = 20.49$ og observert $t = \frac{68.52-44}{20.49/3} = 3.59$, så også nå blir H_0 forkastet

Illustrasjon av t-tester

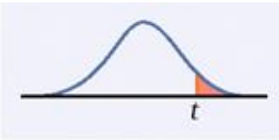
THE ONE-SAMPLE *t* TEST

Suppose that an SRS of size n is drawn from a population having unknown mean μ. To test the hypothesis $H_0: \mu = \mu_0$ based on an SRS of size n, compute the one-sample t statistic

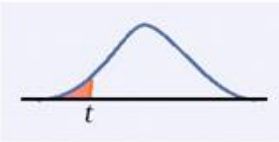
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

In terms of a random variable T having the $t(n-1)$ distribution, the P-value for a test of H_0 against

$H_a: \mu > \mu_0$ is $P(T \geq t)$



$H_a: \mu < \mu_0$ is $P(T \leq t)$



$H_a: \mu \neq \mu_0$ is $2P(T \geq |t|)$



These P-values are exact if the population distribution is normal and are approximately correct for large n in other cases.

Figur fra Moore, McCabe & Craig, brukes i STK1000.

P-verdier, Avsnitt 9.4

I eksemplene foran med tosidig alternativ forkastet vi hvis $|Z| = \left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right|$ eller $|T| = \left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right|$ var stor nok.

Hvis alternativet hadde vært $H_a : \mu > \mu_0$ ville vi forkastet med $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ eller $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ stor nok

og med alternativ $H_a : \mu < \mu_0$ ville vi forkastet med $-Z = -\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ eller $-T = -\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ stor nok

I alle eksempler vi skal møte i STK1110 vil vi kunne uttrykke testen slik at vi forkaster hvis en testobservator $V = V(X_1, \ldots, X_n)$ er stor nok.

Vi kan dermed snakke om et mer ekstremt utfall enn observert verdi $v = v(x_1, \ldots, x_n)$ av V samt om sannsynligheten for dette i en ny tilsvarende studie gitt at nullhypotesen holder.

P-verdier, Avsnitt 9.4, definisjon

Vi antar da at $V = V(X_1, \ldots, X_n)$ er testobservator for valgt nullhypotese H_0 slik at jo større verdi av V jo mindre rimelig er H_0. Vi definerer P-verdien, med observert $V = v$, som

$$P(V \geq v | H_0)$$

der vi nå oppfatter V som en ny verdi av testobservatoren med nye data.

Eksempel 9.12, forts.

- Vi fant testobservator $V = |T| = \left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right|$ for $H_0 : \mu = \mu_0$ som ble forkastet for stor nok V der $T \sim t_{n-1}$ under H_0.
- Observert verdi av V ble $v = 3.59$
- P-verdi for testen blir dermed $P(|T| \geq 3.59 | \mu = \mu_0 = 44)$
  $= P(T \leq -3.59 | \mu = 44) + P(T \geq +3.59 | \mu = 44)$
  $= 2 * 0.00354 \approx 0.0071$ når $T \sim t_8$-fordelt

Konklusjon på hypotest basert på P-verdi

Så sant P-verdien er mindre enn eller lik signifikansnivået α har vi at testobservatoren er ekstrem nok til å forkaster H_0.

På den annen side hvis P-verdien er større enn α forkaster vi ikke, dvs. beholder H_0.

Med P-verdier kan vi angi grad av evidens mot nullhypotesen.
Jo mindre P-verdi desto dårligere overenstemmelse mellom H_0 og data.

P-verdien kan da alternativt defineres som det minste nivå α som gir forkastning.

Hva P-verdien er - og ikke er

P-verdien er altså

- sannsynlighet for et mer ekstremt utfall enn det observerte
- i en ny tilsvarende studie
- gitt at nullhypotesen er sann

P-verdien er imidlertid ikke

- sannsynligheten for at H_0 er sann
- sannsynligheten for at resultatet er generert av ren tilfeldighet

Merk også:

- Testen og P-verdien knytter seg direkte til valgt testobservator!

To be aware of: Lately, there have been extensive discussions in scientific journals and media about the misuse of P-values that has developed in some fields – originating from misconceptions of what P-values mean and how they should be used.



NATURE | NEWS

Statisticians issue warning over misuse of *P* values

Policy statement aims to halt missteps in the quest for certainty.

Monya Baker

07 March 2016

The ASA's Statement on p-Values: Context, Process, and Purpose

The American Statistician 2016 Volume 70(2) pp. 129-133

The statement's six principles, many of which address misconceptions and misuse of the p- value, are the following:

1. P-values can indicate how incompatible the data are with a specified statistical model.

2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

4. Proper inference requires full reporting and transparency.

5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

Avsnitt 9.3: Hypotesetesting om en populasjonsandel p

- Anta at vi har et utvalg $X_1, \ldots, X_n$ fra en stor populasjon, der en andel p har en viss egenskap, med $X_i = 1$ hvis individ/element i har egenskapen og $X_i = 0$ ellers.
- Videre antar vi at $X_1, \ldots, X_n \overset{uif}{\sim} Bernoulli(p)$.
- Da er $Y = \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$.
- Vi er interessert i hypoteser av typen
  - $H_0 : p \leq p_0$ mot $H_a : p > p_0$
  - $H_0 : p \geq p_0$ mot $H_a : p < p_0$
  - $H_0 : p = p_0$ mot $H_a : p \neq p_0$.

Eksempel Salve for eksem, se slide 3.

- 100 individer, eksem begge hender, behandles med ny og gammel salve
- Trekker om høyre eller venstre hånd får ny salve
- $Y =$ antall som responderer best på ny salve $\sim \text{Bin}(n, p)$
- Ny salve best hvis $p > 0.5 = p_0$

Store utvalg

- Anta først at utvalgsstørrelsen n er stor.
- Vi vet at $\hat{p} = \frac{Y}{n}$ er forventningsrett for p og at $\mathrm{V}(\hat{p}) = \frac{p(1-p)}{n}$.
- Hvis $np_0 \geq 10$ og $n(1-p_0) \geq 10$, så er

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \overset{tiln.}{\sim} N(0,1)$$

når $p = p_0$.

- En test med tilnærmet signifikansnivå α for å teste $H_0 : p \leq p_0$ mot $H_a : p > p_0$ forkaster da H_0 dersom

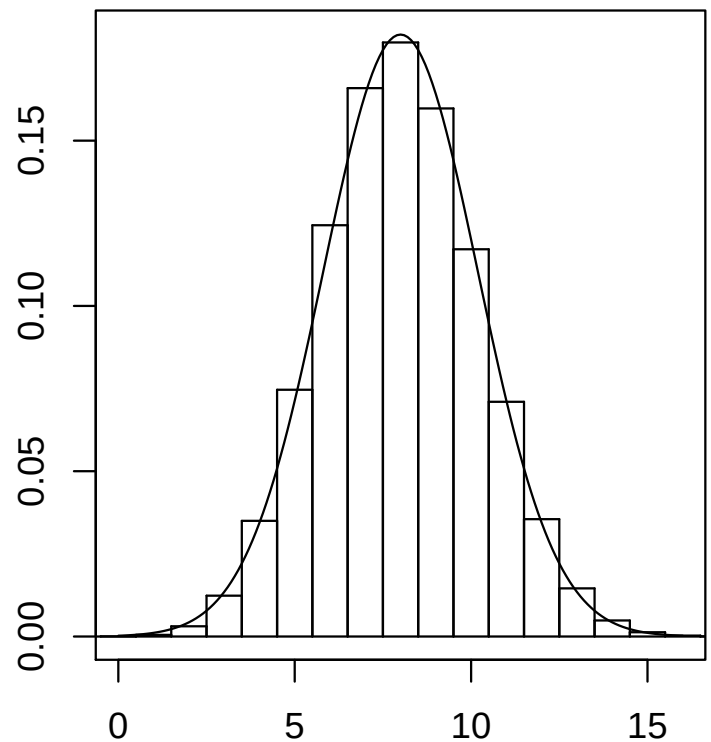
$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \geq z_\alpha.$$

- Eks Salve: $Y \sim \mathrm{Bin}(100, p), \mathrm{H}_0 : p \leq p_0 = 0.5$, $\mathrm{H}_a : p > p_0$
- Observerer (f.eks.) $Y = 60$ som gir $Z = 2.00 > z_{0.05} = 1.645$, forkaster på 5% nivå.
- P-verdi $\mathrm{P}(Z > 2.00 | Z \sim \mathrm{N}(0,1)) = 0.0227$, 2.3%

Halvkorreksjon også for tester om p

Tilsvarende som ved konfidensintervaller bør vi benytte at tilnærming av diskret Y til kontinuerlig fordeling forbedres ved å "smøre" punktsannsynlighetene $\mathrm{P}(Y = y)$ utover intervallet $(y - \frac{1}{2}, y + \frac{1}{2})$, derav blir *halvkorreksjon*

$\mathrm{P}(Y \leq y) \approx \Phi(\frac{y+0.5-np}{\sqrt{np1-p}})$ bedre enn bare $\mathrm{P}(Y \leq y) \approx \Phi(\frac{y-np}{\sqrt{np1-p}})$



Halvkorreksjon for binomiske tester

Vi gir ikke formlene for halvkorrigerte testobservatorer eller konfidensintervall siden disse er en smule kompliserte.

Men halvkorreksjon er implementert i `R`-funksjonen `prop.test` som standard.

Det anbefales å benytte denne framfor den mer intuitive formelen $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ direkte.

Med eksempelet $Y = 60, n = 100$ og $H_a : p > p_0 = .5$ ble p-verdien justert til 0.0287 (fra 0.0227, se `R`-script).

Styrkefunksjon ved store utvalg

Styrkefunksjonen for testen på forrige slide kan uttrykkes

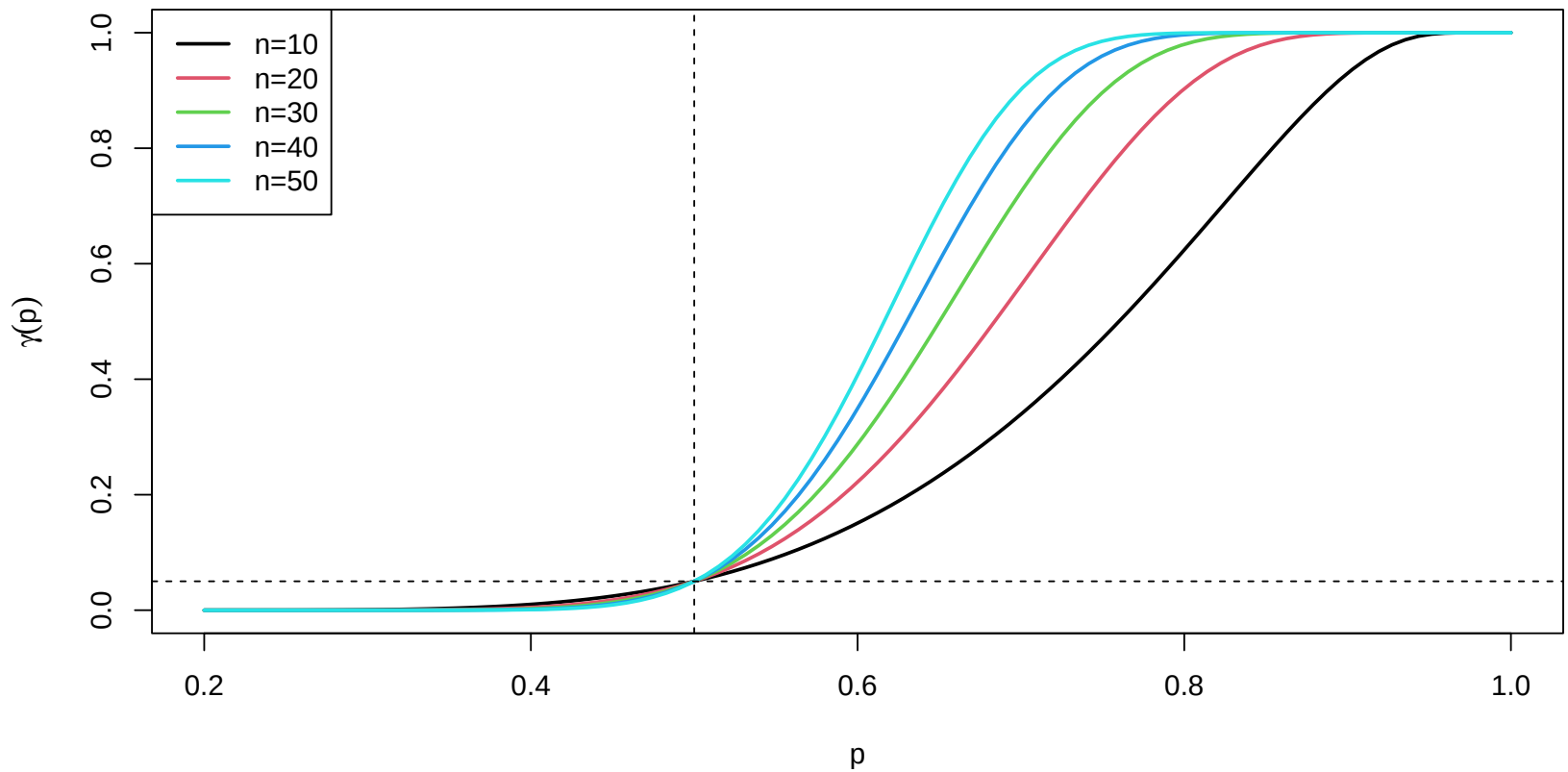
$$\gamma(p) = \mathrm{P}(\text{ Forkaste } H_0|p) \approx 1 - \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$

Utledning:

$$\begin{aligned}
\gamma(p) &= \mathrm{P}(\text{ Forkaste } H_0|p) = \mathrm{P}\left(\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} > z_\alpha \mid p\right) \\
&= \mathrm{P}(\hat{p} > p_0 + z_\alpha\sqrt{p_0(1-p_0)/n} \mid p) \\
&= \mathrm{P}\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} > \frac{p_0-p+z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}} \mid p\right) \\
&\approx 1 - \Phi\left(\frac{p_0-p+z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)
\end{aligned}$$

Illustrasjon styrkefunksjonen ulike n når $H_0 = p \leq 0.5$:

$$\gamma(p) = \mathrm{P}(\text{Forkaste}H_0|p) \approx 1 - \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$



Bestemmelse av n

$$\gamma(p) = \mathrm{P}(\text{Forkaste}H_0|p) \approx 1 - \Phi\left(\frac{p_0 - p + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$

- Vi ønsker å finne n slik at sannsynligheten for feil av type II blir høyst β når $p = p_1 > p_0$.
- Med

$$\begin{aligned}\mathrm{P}(\text{Feil av type II}|p = p_1) =& 1 - \gamma(p_1)\\ \approx& \Phi\left(\frac{p_0 - p_1 + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p_1(1-p_1)/n}}\right) \leq \beta\end{aligned}$$

får vi

$$n \geq \left(\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_1 - p_0}\right)^2.$$

Styrkefunksjon ved store utvalg (forts.)

- For tester av typen $H_0 : p \geq p_0$ mot $H_a : p < p_0$ forkaster en H_0 ved signifikansnivå tilnærmet α dersom $Z \leq -z_\alpha$.
- Styrkefunksjonen er da

$$\gamma(p) \approx \Phi\left(\frac{p_0 - p - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$

- Den n som gir sannsynlighet for feil av type II høyst lik β når $p = p_1 < p_0$ er da som på forrige foil.

Styrkefunksjon, tosidig

- For tester av typen $H_0 : p = p_0$ mot $H_a : p \neq p_0$ forkaster en H_0 ved signifikansnivå tilnærmet α dersom $Z \leq -z_{\alpha/2}$ eller $Z \geq z_{\alpha/2}$.
- Styrkefunksjonen er da

$$\gamma(p) \approx 1 - \Phi\left(\frac{p_0 - p + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) + \Phi\left(\frac{p_0 - p - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right).$$

- Den n som gir sannsynlighet for feil av type II høyst lik β når $p = p_1 \neq p_0$ er da tilnærmet

$$n \geq \left(\frac{z_{\alpha/2}\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_1 - p_0}\right)^2.$$

Små utvalg

- Dersom n ikke er stor nok til at vi kan bruke tilnærmingen med normalfordeling, kan vi i stedet bruke den binomiske fordelingen direkte.
- Anta at vi vil teste $H_0 : p \leq p_0$ mot $H_a : p > p_0$.
- En forkaster da H_0 for $Y \geq k$ for en passende k.
- Vi har

$$\mathrm{P}(\text{Feil av type I}) \leq 1 - B(k-1; n, p_0),$$

der $B(\cdot; n, p_0)$ er den kumulative fordelingsfunksjonen til den binomiske fordelingen med parametere (n, p_0).

Små utvalg (forts.)

- Det er sjelden mulig å finne en k slik at sannsynligheten for feil av type I blir akkurat lik signifikansnivået.
- I stedet velger en det minste heltallet k slik at

$$1 - B(k-1; n, p_0) \leq \alpha.$$

- For $p_1 > p_0$ får vi sannsynlighet for feil av type II lik $\beta(p_1) = B(k-1; n, p_1)$.
- Tester av typen $H_0 : p \geq p_0$ mot $H_a : p < p_0$ finner en på samme måte.
- En forkaster da H_0 ved signifikansnivå α når $Y \leq k$, der k er det største heltallet slik at

$$\text{P(Feil av type I)} \leq B(k; n, p_0) \leq \alpha.$$

Små utvalg (forst.)

- For tester av typen $H_0 : p = p_0$ mot $H_a : p \neq p_0$ forkaster en H_0 ved signifikansnivå α når $Y \leq k_1$ eller $Y \geq k_2$, med $k_1 < k_2$, der de utgjør det minste forkastningsområdet slik at

$$\text{P(Feil av type I)} \leq B(k_1; n, p_0) + 1 - B(k_2 - 1; n, p_0) \leq \alpha.$$

dvs. slik at

$$B(k_2 - 1; n, p_0) - B(k_1 - 1; n, p_0) \geq 1 - \alpha.$$

Eksakt binomisk test i R

Man kan benytte implementasjonen av kumulative for binomiske data `pbinom` direkte.

Men det er nok enklere å heller bruke den eksakte testen i `binom.test`, ikke minst siden denne også gir eksakte konfidensintervall.

For eksempelet med $Y = 60, n = 100$ og $H_a : p > 0.5$ finner vi p-verdien 0.0284, altså temmelig likt som den halvkorrigerte testen (p=0.0287)!

Avsnitt 9.5: Likelihood-ratio tester (s. 547-551)

- Så langt: Test konstruert vha en passende test-observator
  - Typisk fra en estimator for parameter av interesse
- Trenger noen ganger en mer generell prosedyre:
  - Likelihood ratio tester

Likelihood-ratio-tester

- Anta at $X_1, \ldots, X_n$ er stokastiske variabler med simultantetthet/-punktsannsynlighet $f(x_1, \ldots, x_n; \theta)$, der θ kan være en vektor, dvs. $\theta = (\theta_1, \ldots, \theta_k)$.
- Parameterrommet Ω er mengden av mulige verdier for θ, dvs. $\theta \in \Omega$.
- La $\Omega_0 \subset \Omega$ og $\Omega_a = \Omega \setminus \Omega_0$.
- Vi vil teste hypotesen $H_0 : \theta \in \Omega_0$ mot alternativet $H_a : \theta \in \Omega_a$.
  - Eksempel 1: $Y \sim \text{Bin}(n, p)$, $H_0 : p \leq p_0$
    - $\Omega = [0, 1], \Omega_0 = [0, p_0]$
  - Eksempel 2: $X_i \sim N(\mu, \sigma^2)$, $H_0 = \mu \leq \mu_0$:
    - $\Omega = \{(\mu, \sigma); -\infty < \mu < \infty, 0 < \sigma < \infty\}$
    - $\Omega_0 = \{(\mu, \sigma); \mu \leq \mu_0, 0 < \sigma < \infty\}$

Likelihood-ratio-tester (forts.)

- La $x_1, \ldots, x_n$ være observerte verdier av $X_1, \ldots, X_n$.
- Da er likelihood-funksjonen gitt ved $L(\theta) = f(x_1, \ldots, x_n; \theta)$.
- La nå $\hat{\theta}$ være MLE for θ.
- Videre, la $\hat{\theta}_0$ være MLE for θ under H_0, altså den verdien av θ som maksimerer $L(\theta)$ under begrensningen $\theta \in \Omega_0$.
- Da er likelihood-ratioen (sannsynlighetskvoten) gitt ved:

$$LR = \frac{\max_{\theta \in \Omega_0} L(\theta)}{\max_{\theta \in \Omega} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}.$$

Likelihood-ratio-tester (forts.)

- Da $\Omega_0 \subset \Omega$, må nødvendigvis $L(\hat{\theta}) \geq L(\hat{\theta}_0)$, slik at $LR \leq 1$.
- Dersom LR er liten, betyr det at dataene samsvarer mye mer med H_a enn med H_0.
- Dermed forkaster vi H_0 dersom $LR \leq k$, der k velges slik at vi får ønsket signifikansnivå α.

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$, der σ er kjent. Vi vil teste $H_0 : \mu = \mu_0$ mot $H_a : \mu \neq \mu_0$.

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$, der σ er ukjent. Vi vil teste $H_0 : \mu = \mu_0$ mot $H_a : \mu \neq \mu_0$.

Eks. $X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$, der σ er kjent.

Vi vil teste $H_0 : \mu = \mu_0$ mot $H_a : \mu \neq \mu_0$

Under Ω blir MLE $\hat{\mu} = \bar{X}$ og under $\Omega_0 = \{\mu_0\}$ blir MLE $\hat{\mu}_0 = \mu_0$.
Dermed får vi

$$LR = \frac{L(\mu_0)}{L(\bar{X})} = \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu_0)^2)}{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\bar{X})^2)}$$

$$= \frac{\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu_0)^2)}{\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\bar{X})^2)}$$

Dette gir videre at

$$-2\ln(LR) = \frac{1}{\sigma^2}\left[\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\bar{X})^2\right] = \cdots = \frac{n(\bar{X}-\mu_0)^2}{\sigma^2}$$

$$= Z^2$$

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

Eks. $X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma)$, der σ er kjent.

Vi har dermed at LR liten nok til å forkaste er det samme som

$$-2\ln(LR) = Z^2 = \frac{n(\bar{X}-\mu_0)^2}{\sigma^2} \text{ stor nok til å forkaste og}$$

LR-testen er ekvivalent med vanlig test basert på Z^2 eller $|Z|$.

Legg videre merke til at her har vi *eksakt*

$$-2\ln(LR) \sim \chi^2_1$$

under $H_0 : \mu = \mu_0$.

Denne egenskapen gjelder generelt *tilnærmet* for $-2\ln(LR)$ (faktisk) når forskjellen i dimensjon mellom Ω og Ω_0 er lik 1.

Eks. $X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$, der σ er ukjent

og fortsatt $H_0 : \mu = \mu_0$

Under Ω er MLE for μ, σ^2 gitt ved
$(\bar{X}, \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2) = (\hat{\mu}, \hat{\sigma}^2)$.

Under $\Omega_0 = \{\mu_0\} \times (0, \infty)$ blir MLE for (μ, σ^2) gitt ved
$(\mu_0, \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2) = (\mu_0, \hat{\sigma}_0^2)$.

Vi får da

$$LR = \frac{L(\mu_0, \hat{\sigma}_0^2)}{L(\bar{X}, \hat{\sigma}^2)} = \frac{\frac{1}{(\hat{\sigma}_0^2)^{n/2}} \exp(-\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (X_i - \mu_0)^2)}{\frac{1}{(\hat{\sigma}^2)^{n/2}} \exp(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n (X_i - \bar{X})^2)} = \left[\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right]^{n/2}$$

siden $\frac{1}{2\hat{\sigma}_0^2}\sum_{i=1}^n (X_i - \mu_0)^2 = \frac{n}{2} = \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^n (X_i - \bar{X})^2$

Dermed får vi at en liten LR er det samme som en liten verdi av

$$\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \mu_0)^2}$$

Eks. $X_1, \ldots, X_n \overset{uif}{\sim} N(\mu, \sigma^2)$, der σ er ukjent, forts.

Videre kan vi skrive

$$\sum_{i=1}^{n}(X_i-\mu_0)^2 = \sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu_0)^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2+n(\bar{X}-\mu_0)^2$$

Dermed blir liten LR og liten $\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}$ det samme som stor

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2+n(\bar{X}-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = 1+\frac{n(\bar{X}-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

som igjen er ekvivalent med stor

$$T^2 = \frac{n(\bar{X}-\mu_0)^2}{S^2}$$

der S^2 er empirisk varians og $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ under $H_0 : \mu = \mu_0$.

Mao. LR testen og T-testen er også ekvivalente!

Likelihood-ratio-tester for store utvalg

- Anta at $X_1, \ldots, X_n$ er uif med tetthet/punktsannsynlighet $f(x;\theta)$, som avhenger av én parameter θ.
- Vi vil teste $H_0 : \theta = \theta_0$ mot $H_a : \theta \neq \theta_0$.
- Hvis n er stor, er $-2\ln(LR) \overset{tiln.}{\sim} \chi^2_1$ når H_0 er sann.
- Vi får dermed en test med tilnærmet signifikansnivå α hvis vi forkaster H_0 så sant $-2\ln(LR) \geq \chi^2_{\alpha,1}$.

Eksempel

$X \sim Bin(n, p)$. Vi vil teste $H_0 : p = p_0$ mot $H_a : p \neq p_0$.

LR-test for $H_0 : p = p_0$ mot $H_a : p \neq p_0$

med $X \sim \text{Bin}(n, p)$.

Vi har $\hat{p} = \frac{X}{n}$ MLE under $\Omega = (0, 1)$ og får dermed

$$LR = \frac{L(p_0)}{L(\hat{p})} = \left[\frac{p_0}{\hat{p}}\right]^X \left[\frac{1-p_0}{1-\hat{p}}\right]^{n-X}$$

som gir

$$-2\ln(LR) = -2X\ln(\frac{p_0}{\hat{p}}) - 2(n-X)\ln(\frac{1-p_0}{1-\hat{p}}) \sim \chi^2_1 \quad \text{(tilnærmet)}$$

Vi fant tidligere at for $n = 100, X = 60$ og $p_0 = 0.5$ ble $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}} = 2$ og dermed $Z^2 = 4$, samtidig som $Z^2 \sim \chi^2_1$ tilnærmet

Innsatt $n = 100, X = 60$ og $p_0 = 0.5$ blir $-2\ln(LR) = 4.027$
Kan vise tilnærmelsen mellom $-2\ln(LR)$ og Z^2 ved Taylor-utvikl.

MLE i dobbel eksponensiell fordeling

Anta $X_i \sim f(x;\theta) = \frac{1}{2}\exp(-|x-\theta|)$ for $x \in \Re$, $i = 1, \ldots, n$, uif.

Dermed blir log-likelihood

$$\ell(\theta) = -n\ln(2) - \sum_{i=1}^{n} |X_i - \theta|$$

og maksimering av $\ell(\theta)$ er ekvivalent med minimering av sum av absolutt-avvik $\sum_{i=1}^{n} |X_i - \theta|$.

Dette har faktisk løsningen MLE $\hat{\theta}$ = medianen til $X_1, \ldots, X_n$ (se læreboka, s. 551)

Riktignok er ikke $\ell(\theta)$ deriverbar i observerte X_i. Likevel holder large sample teorien for dette tilfellet og

$$-2\ln(LR) = 2(\sum_{i=1}^{n} |X_i - \theta_0| - \sum_{i=1}^{n} |X_i - \hat{\theta}|) \sim \chi^2_1$$

tilnærmet for test av $H_0 : \theta = \theta_0$ mot $H_a : \theta \neq \theta_0$.

Likelihood-ratio-tester for store utvalg (forts.)

- Anta at $X_1, \ldots, X_n$ er uif med tetthet/punktsannsynlighet $f(x; \theta_1, \ldots, \theta_k)$, som avhenger av k parametere $\theta_1, \ldots, \theta_k$.
- Vi vil teste $H_0 : (\theta_1, \ldots, \theta_k) \subset \Omega_0$ mot $H_a : (\theta_1, \ldots, \theta_k) \subset \Omega_a$.
- Videre antar vi at hvis H_0 er sann, har vi $m < k$ parametere.
- Hvis n er stor, er da $-2\ln(LR) \overset{tiln.}{\sim} \chi^2_{k-m}$ når H_0 er sann.

Eksempel

Er nedbørsdata eksponensialfordelt eller gammafordelt?

Testing av om eksponensialfordeling passer

Anta at $X_i \sim f(x;\alpha,\beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}\exp(-\frac{x}{\beta})$ for $x > 0$, dvs. gammafordelt, $i = 1,\ldots,n$ uif.

Hvis $H_0 : \alpha = 1 = \alpha_0$ er X_i eksponensialfordelt med $\mathrm{E}(X_i) = \beta$. Vi kan derfor være interessert i å teste H_0 mot $H_a : \alpha \neq \alpha_0 = 1$.

Her blir $\Omega = \{(\alpha,\beta) : \alpha > 0 \text{ og } \beta > 0\}$ og $\Omega_0 = \{(1,\beta) : \beta > 0\}$

Samtidig får vi tilnærmet med stor n

$$-2\ln(LR) = -2(\ell(\hat\alpha,\hat\beta) - \ell(1,\beta^*)) \sim \chi^2_1$$

når $(\hat\alpha,\hat\beta)$ er MLE over Ω, mens $\beta^* = \bar X$ er MLE for β under H_0 (og over Ω_0).

MLE $(\hat\alpha,\hat\beta)$ må finnes ved numerisk optimering av

$$\begin{aligned}\ell(\alpha,\beta) &= \textstyle\sum_{i=1}^n \ln(f(X_i;\alpha,\beta)) \\ &= \textstyle\sum_{i=1}^n \left(-\alpha\ln(\beta) - \ln(\Gamma(\alpha)) + (\alpha-1)\ln(X_i) - \frac{X_i}{\beta}\right)\end{aligned}$$

P-verdi som stokastisk variabel

Vi definerte P-verdien (slide 8) basert på testobservator $V = V(X_1, \ldots, X_n)$ for valgt nullhypotese H_0 slik at jo større verdi av V jo mindre rimelig er H_0.
Spesifikt er P-verdien, med obs. $V = v$,

$$\mathrm{P}(V_{ny} \geq v | H_0)$$

der V_{ny} er en ny verdi av testobservatoren med nye data.

Men den faktiske testobservatoren er jo basert på opprinnelige stokastiske variable $(X_1, \ldots, X_n)$ og dermed er P-verdien en funksjon av disse og dermed selv en stokastisk variabel med en fordeling.

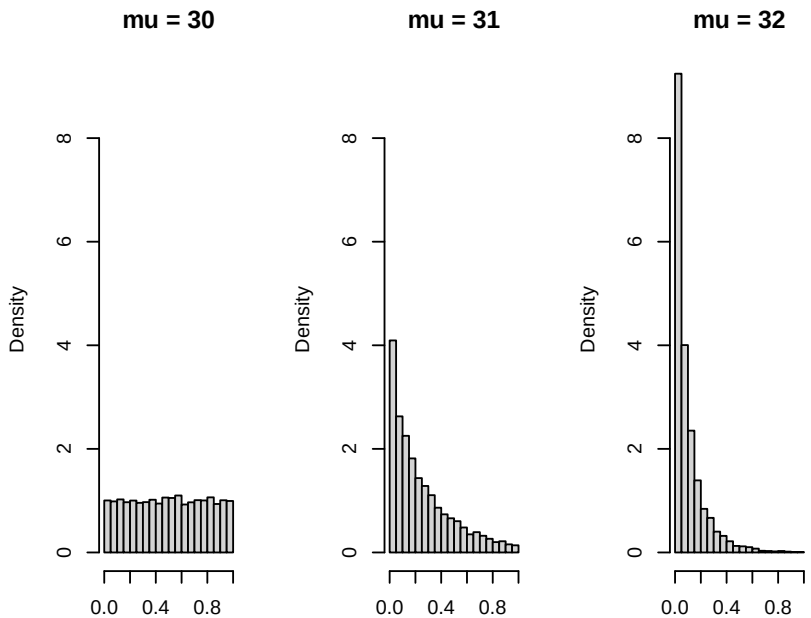
Kan vises at hvis V har kontinuerlig fordeling så vil P-verdien som stokastisk variabel $W = \mathrm{P}(V_{ny} \geq V(X_1, \ldots, X_n) | H_0)$

- ha en fordeling på intervallet $[0, 1]$
- være uniformt fordelt hvis H_0 er sann
- være skjev mot små verdier når H_a er sann

Eksempel P-verdi som stokastisk variabel, Eks. 9.22

Vil studere fordelingen til P-verdien ved simulering.

- Situasjon: $X_1, \ldots, X_n$ uif $\mathrm{N}(\mu, \sigma)$ med $n = 4$ og σ ukjent.
- $H_0 : \mu = \mu_0 = 30$ mot $H_a : \mu > 30$ med $\alpha = 0.05$.
- Så testobservator blir $T = \frac{\bar{X} - 30}{S/\sqrt{n}} \sim t_3$ under H_0.
- P-verdien blir da $W = P(t_3 > \frac{\bar{X} - 30}{S/\sqrt{n}} \mid \mu)$
- Trekker X_1, X_2, X_3, X_4 fra $\mathrm{N}(\mu, 2)$ og beregner P-verdi
- N=10000 ganger
- med ulike verdier av $\mu = 30 (H_0), \mu = 31$ og $\mu = 32$



Statistisk signifikans og praktisk relevans

- Merk at det å ikke forkaste (altså beholde) H_0 ikke er det samme som å ha vist (selv med rimelig grad av sikkerhet) at H_0 er sann.
- Det kan tenkes at H_0 ikke er sann, men at dataene ikke har nok "beviskraft" til å vise det, f.eks. at utvalget er for lite.
- Videre er et **statistisk signifikant** resultat ikke nødvendigvis det samme som et resultat av **praktisk relevans**.
- Når n er stor, er det stor sannsynlighet for å forkaste H_0, altså få et statistisk signifikant resultat, selv om forskjellen mellom den sanne θ og nullverdien θ_0 er så liten at den ikke er relevant i praksis.
- Fortolkning av P-verdier og betydningen av statistisk signifikans har vært mye diskutert de siste årene, se f. eks. https://forskning.no/matematikk-om-forskning-samfunn/hva-er-p-verdi-og-hva-betyr-statistisk-signifikant/1321080

  (Artikkelen på forskning.no)