Mathematical Statistics II

STA2212H S LEC0101

Week 6

February 14 2023


Nordic Summer Evening Richard Bergh
Daily Art App

Today

1. Recap: exact and approx CIs; credible intervals; Bayes asymptotics; confidence bands
2. Confidence and HPD regions
3. Hypothesis testing
4. Project

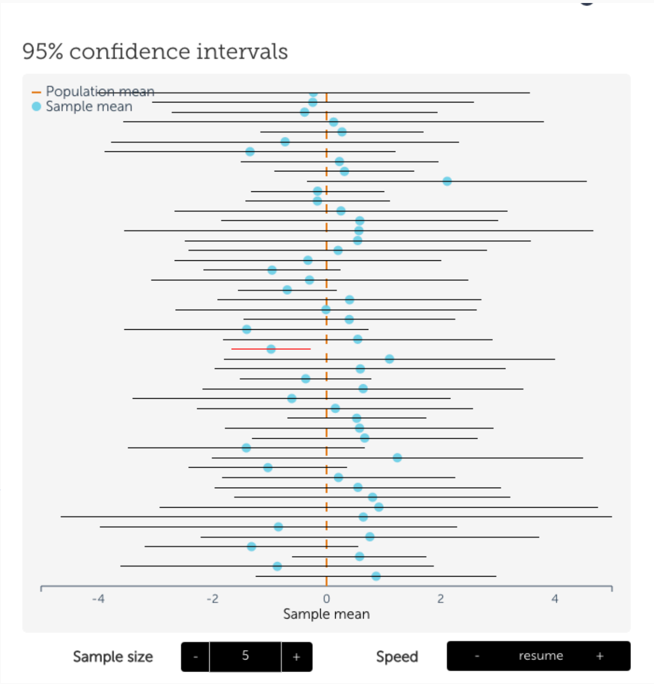
Upcoming

- Reading week! — No office hours
- February 22-23 Toronto Workshop on Reproducibility

Recap

- frequentist interpretation of CIs
- exact and approximate confidence intervals
- exact and approximate credible intervals

visualization



Est^{n} CIs Testing hyp.

exact approx ex. approx exact approx

good $n < \infty$

good enough

pivotal exact dist

convent. approx =

most powerful optimal $n < \infty$

lik-based (same)

(unbiased) min. var. admissible minimax opt. Bayes

$\hat{\theta}_{MLE}$

pivotal

Bayes credible intervals

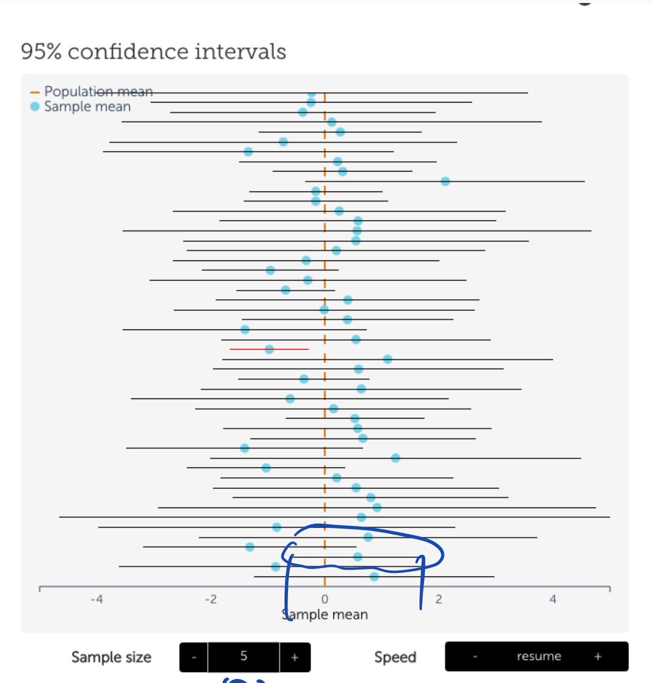
good enough (Bayes tests)

Recap

- frequentist interpretation of CIs
- exact and approximate confidence intervals
- exact and approximate credible intervals

- asymptotic theory: likelihood and posterior

visualization



$$\int_{a(x)}^{b(x)} \pi(\theta \mid x)\, d\theta = 1-\alpha$$

$\Rightarrow$ θ

$$\Pr_{\pi(\theta \mid x)}\{ a(x) \le \theta \le b(x) \mid x \} = 1-\alpha$$

– for $n \to \infty$ $\pi(\theta)$ is "washed out" by data if $p < \infty$

dep. on $\pi(\theta)$

Bayes interp. $= \Pr_{\pi(\theta \mid x)}(-1 \le \theta < 1.5) = .95$

Recap

- for $n < \infty$ $\pi(\theta)$ might have large effect if p is large
- if $p = p_n \uparrow$ w/ n → asy. theory fails

- frequentist interpretation of CIs
- exact and approximate confidence intervals
- exact and approximate credible intervals

visualization

$E(y|x) = m(x)$
smooth fn.

- asymptotic theory: likelihood and posterior

AoS ← an approx. interval

- confidence bands: pointwise or simultaneous

$$pr\{a(\underset{\sim}{X}) \le m(x) \le b(\underset{\sim}{X})\} \approx 0.95 \quad (1-\alpha)$$

for given x.

$$pr\{a(\underset{\sim}{X}) \le m(x) \le b(\underset{\sim}{X}) \ \forall x\} \ge .95$$

$1-\alpha$

324 20. Nonparametric Curve Estimation

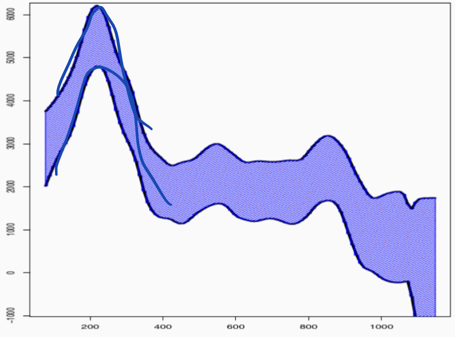


FIGURE 20.9. 95 percent confidence envelope for the CMB data.

Vector parameters: likelihood ratio confidence regions

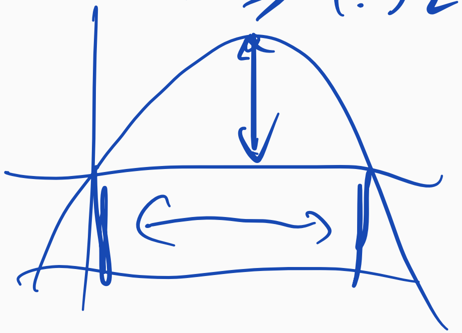
- $X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$ $\qquad \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$

- $L(\boldsymbol{\theta}; \boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta}), \quad \ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}; \boldsymbol{x})$

- $$w(\boldsymbol{\theta}) = 2\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\} \xrightarrow{d} \chi^2_p, \quad n \to \infty$$

$\ell(\theta_1, \theta_2; \underset{\sim}{x})$

$w \geq 1.92$



Vector parameters: likelihood ratio confidence regions

- $X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$
- $L(\theta; \boldsymbol{x}) = f(\boldsymbol{x}; \theta), \quad \ell(\theta) = \log L(\theta; \boldsymbol{x})$
- $$w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\} \xrightarrow{d} \chi^2_p, \quad n \to \infty$$
- approximation:
$$w(\theta) \dot{\sim} \chi^2_p$$
- approximate confidence region
$$\{\theta : w(\theta) \geq \chi^2_{p,1-\alpha}\}$$

1 - α conf. region

Highest posterior density (HPD) regions

$\alpha/2$ $\alpha/2$

- HPD region C for θ:

$\theta \in \mathbb{R}$

$2\{L(\hat{\theta};x) - L(\theta;x)\} \geq \chi^2_{1-\alpha}$

$$(1) \quad \int_C \pi(\theta \mid \boldsymbol{x}) = 1 - \alpha$$

$$(2) \quad \pi(\theta \mid \boldsymbol{x}) \geq \pi(\theta^* \mid \boldsymbol{x}) \qquad \theta^* \notin C$$

Binomials

A B . . . M

0/45

580 — *11 · Bayesian Models*

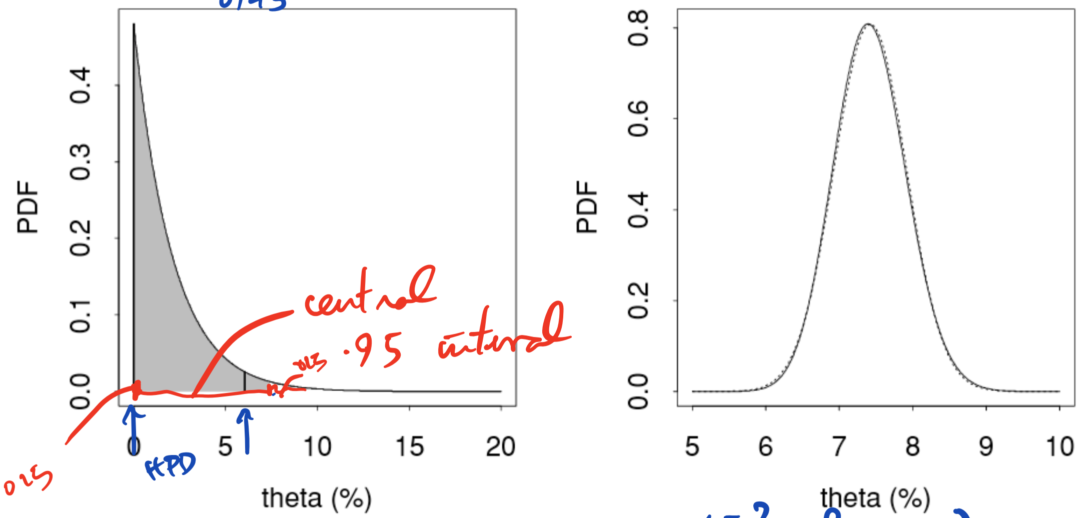


Figure 11.1 Cardiac surgery data. Left panel: posterior density for θ_A, showing boundaries of 0.95 highest posterior credible interval (vertical lines) and region between posterior 0.025 and 0.975 quantiles of $\pi(\theta_A \mid y)$ (shaded). Right panel: exact posterior beta density for overall mortality rate θ (solid) and normal approximation (dots).

$\alpha = ?,\ \beta = ?)$

582 *11 · Bayesian Models*

(μ,σ^2) $1-\alpha$ $\pi(\theta|x)$

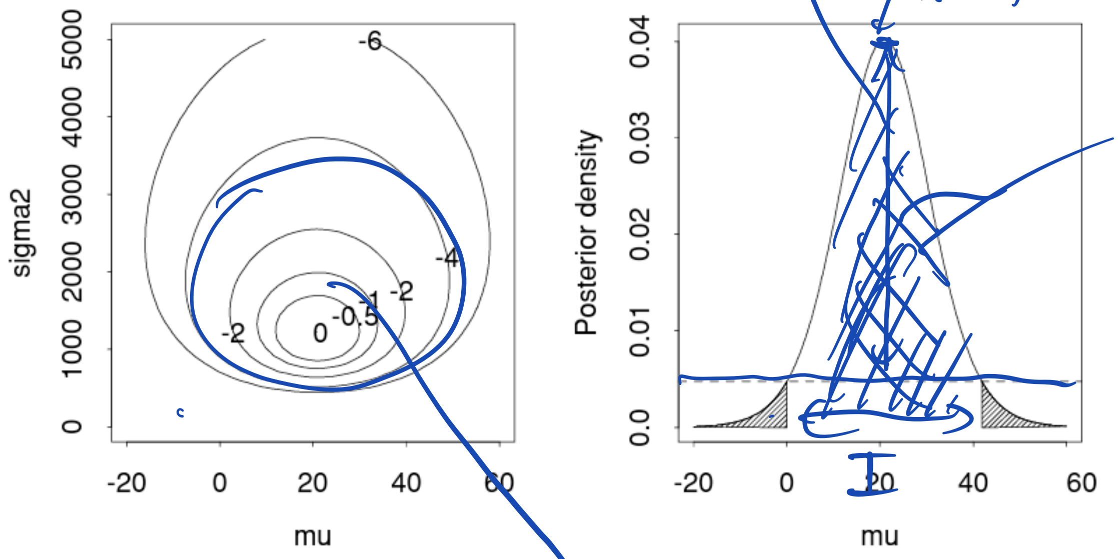


Figure 11.2 Posterior densities of (μ,σ^2) of normal model for maize data. Left: contours of the normalized log joint posterior density. Right: marginal posterior density for μ, showing 95% HPD credible set, which is the set of values of μ whose values of the posterior density $\pi(\mu \mid y)$ lie above the dashed line. The shaded region has area 0.05.

$\theta \in I$

$\pi(\theta|x) \geq$

$\pi(\theta^*|x)$

$\theta^* \in I$

$\int_C \pi(\theta|x)\,d\theta = 1-\alpha$

$N(\mu,\sigma^2)$

$x_1,\dots,x_n$ iid

$\pi(\theta|\underset{\sim}{x}) \geq \ell$

Introduction to Formal Hypothesis testing

$(\Theta_0 \cup \Theta_1 = \Theta)$

$X_1, \ldots, X_n \overset{iid}{\sim} f(\mathbf{x}; \theta)$ $\theta \in \Theta$

- Null and alternative hypothesis $H_0: \theta \in \Theta_0 \subset \Theta$ $H_1: \theta \in \Theta_1$ (sometimes) $\theta \notin \Theta_0$
- Test function
- Rejection region
- Type I and Type II error
- Power and Size

$X_1, \ldots X_n$ $N(\mu_x, \sigma^2)$

$Y_1, \ldots Y_m$ $N(\mu_y, \sigma^2)$

$\Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$

$\Theta_0 : \mu_x = \mu_y$ $\mathbb{R} \times \mathbb{R}^+$

$\phi : \mathcal{X} \to \{0, 1\}$

s-space

$\phi(\underline{x}) = \begin{cases} 1 & \text{reject } H_0 \\ 0 & \text{do not reject } H_0 \end{cases}$

$X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$

sometimes

- Null and alternative hypothesis: $H_0 : \theta \in \Theta_0$; $H_1 : \theta \in \Theta_1$, $\quad \Theta_0 \cup \Theta_1 = \Theta$
- Test (decision) function: $\phi : \mathcal{X} \to \{0, 1\}$
  $\phi(\boldsymbol{X}) = 1$ decide $\theta \cancel{\in} \Theta_0$, else decide $\theta \in \Theta_0$

links to CIs

- Rejection region: $R \subset \mathcal{X}$; if $\boldsymbol{x} \in R$ "reject" H_0 ; $x \notin R$ do not reject $\quad R = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 1\}$
- Type I and Type II error: $\Pr\{\boldsymbol{X} \in R \mid \theta \in \Theta_0\}$, $\quad \Pr\{\boldsymbol{X} \notin R \mid \theta \in \Theta_1\}$

rej. H_0, but it's true

don't reject H_0, but it's false

- Power and Size: $\beta(\theta) = \Pr_\theta(X \in R)$ $\quad \alpha = \sup_{\theta \in \Theta_0} \beta(\theta)$
- Optimal tests: among all level-α tests, find that with the highest power under H_1

level-α means size $\leq \alpha$

fix type I error to be small (α), then minimize type II error

- goal is to identify *R*, or $\phi(\cdot)$ with small Type I and Type II errors
- can't reduce both errors at once

see text following Ex. 7.10

- classical solution: require

$$E_\theta\{\phi(\boldsymbol{X})\} \leq \alpha, \quad \theta \in \Theta_0$$

- subject to this constraint, minimize

$$E_\theta\{\phi(\boldsymbol{X})\}, \quad \theta \in \Theta_1$$

$\phi(\underline{x}) = 1 \Rightarrow$ "reject H_0"

$\Rightarrow P_{\theta\in\Theta_0}(\phi(\underline{X}) = 1)$

$\parallel$

$P_\theta(\phi(x) = 1), \ \theta \in \Theta_1$

max $P_\theta(\phi(\underline{x}) = 0); \ \theta \in \Theta_1$

II

reject H_0 when it's false

- goal is to identify R, or $\phi(\cdot)$ with small Type I and Type II errors
- can't reduce both errors at once

see text following Ex. 7.10

- classical solution: require

$$E_\theta\{\phi(\boldsymbol{X})\} \leq \alpha, \quad \theta \in \Theta_0$$

size of test

- subject to this constraint, minimize

$$\underbrace{E_\theta\{\phi(\boldsymbol{X})\}}_{\text{type 2 error}}, \quad \theta \in \Theta_1 \equiv \max.\{1 - \text{type II error}\}$$

↑ power

- find a test statistic, $T = t(\boldsymbol{X})$, and $\phi(\boldsymbol{X}) = 1\{T \geq t_{crit}\}$

t_{crit} to be determined

larger T is, more evidence we have

against H_0 .

1.2 Hypothesis Testing

Our second example concerns the march of methodology and inference for *hypothesis testing* rather than estimation: 72 leukemia patients, 47 with ALL (acute lymphoblastic leukemia) and 25 with AML (acute myeloid leukemia, a worse prognosis) have each had genetic activity measured for a panel of 7,128 genes. The histograms in Figure 1.4 compare the genetic activities in the two groups for gene 136.

$$\left.\begin{array}{l} X_{1i} \sim N(\mu_1, \sigma^2) \\ X_{2i} \sim N(\mu_2, \sigma^2) \end{array}\right]$$

$H_0: \mu_1 = \mu_2$

$H_0: \mu_1 - \mu_2 = 0$

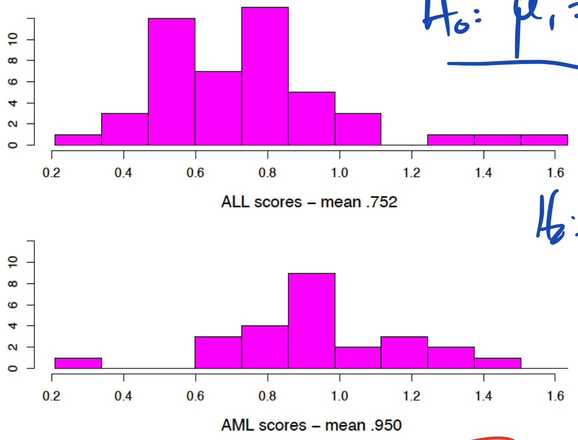
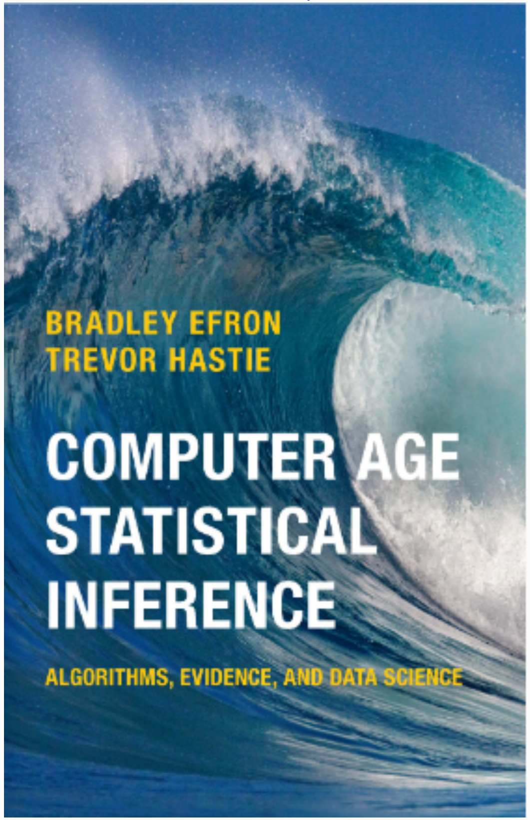


Figure 1.4 Scores for gene 136, leukemia data. Top ALL ($n = 47$), bottom AML ($n = 25$). A two-sample t-statistic $= 3.01$ with p-value $= .0036$.

The AML group appears to show greater activity, the mean values being

$$\overline{\text{ALL}} = 0.752 \quad \text{and} \quad \overline{\text{AML}} = 0.950. \tag{1.5}$$



$n = 72$

$p = 7136$ genes

```
leukemia_big <- read.csv
  ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
oneline <- leukemia_big[136,]
one <- c(1:20, 35:61) #  I had to extract these manually,
two <- c(21:34, 62:72) #  couldn't figure out the data frame
n1 <- length(one); n2 <- length(two)
mean_one <- sum(oneline[1,one])/n1.  ##[1] 0.7524794
mean_two <- sum(oneline[1,two])/n2.  ##[1] 0.9499731
var_one <-  sum((oneline[1,one]-mean_one)^2)/(n1-1)
var_two <- sum((oneline[1,two]-mean_two)^2)/(n2-1)
pooled <- ((n1-1)*var_one + (n2-1)*var_two)/(n1+n2-1)
taos <- (mean_one-mean_two)/sqrt((var_one/n1)+(var_two/n2))
##[1] -3.132304
tbe <- (mean_one-mean_two)/sqrt(pooled*((1/n1)+(1/n2)))
##[1] -3.035455
```

gene 136

ALL

AML

$\bar{x}_1 = 0.7525$

$\bar{x}_2 = 0.9499$

$\sum (x_{1i} - \bar{x}_1)^2 / (n_1 - 1)$

$$S^2 = \frac{\sum (x_{1i} - \bar{x}_1)^2 + \sum (x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

2

1

$$(\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

... t-test

- model $X_{11}, \ldots X_{1n_1}$ iid $N(\mu_1, \sigma^2)$; $X_{21}, \ldots, X_{2n_2}$ iid $N(\mu_2, \sigma^2)$
- null and alternative hypothesis $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
- rejection region
- test statistics and critical value
- type I and type II error

it

$$t(\underset{\sim}{X}_1, \underset{\sim}{X}_2) = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$R = \{ \underset{\sim}{x}_1, \underset{\sim}{x}_2 : t(\underset{\sim}{X}_1, \underset{\sim}{X}_2) \geq t^{\alpha/2}_{n_1+n_2-2} \}$$

$\bar{x}_1 \quad \bar{x}_2 \quad s^2$

$$\text{power} = P_{H_1}(t \geq t^{\alpha/2})$$

$H_1 \equiv \mu_1 \neq \mu_2$

choose a test stat.

figure out it's dist^n

Need dist^n under H_0 for $\leq \alpha$

" " " H_1 to assess power

$$T(\underset{\sim}{x}) = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad |T| > t$$

evidence $\mu_1 - \mu_2 \neq 0$

under H_0 $T(\underset{\sim}{X}_1, \underset{\sim}{X}_2) \sim t_{n_1+n_2-2}$ d.f.

$t^{\alpha/2}_{72-2}$ value is $\boxed{2.0}$

$3.01 > 2.0$

$\phi(\underset{\sim}{x}_1, \underset{\sim}{x}_2) = 1$ bec. $t(\underset{\sim}{x}_1, \underset{\sim}{x}_2) \geq 2.0$

p-value $.002 = Pr_{H_0}\{|t(\underset{\sim}{X}_1, \underset{\sim}{X}_2)| \geq |t^{obs}| = 3.01\}$

Example: comparing two proportions

model, null, alternative, rejection reg, test stat, …

Example: comparing two proportions

Pred. algo 1 ; test set 1

- $X \sim Binom(m, p_1), Y \sim Binom(n, p_2)$ ← prediction algo 2 — test set 2

two prediction algorithms

- $\delta = p_1 - p_2; H_0 : \delta = 0$
- maximum likelihood estimate of δ =
- estimated standard error

$$\frac{\hat{\delta} - 0}{\hat{se}(\hat{\delta})}$$

$$t(X,Y) = \left(\frac{X}{m} - \frac{Y}{n}\right) \Big/ \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}$$

$$\hat{p}_1 - \hat{p}_2 = \hat{\delta}$$

$$\sim N(0,1)$$

$$Pr_{H_0}\{|t(X,Y)| \geq k\}$$

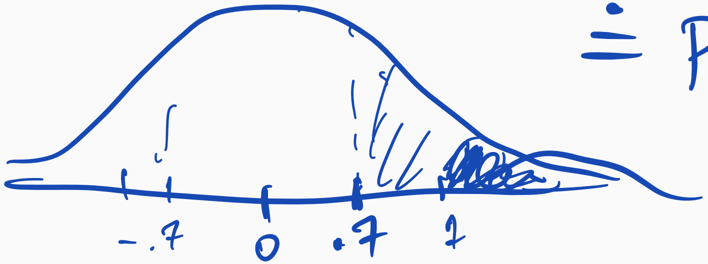
$$\doteq Pr(|N(0,1)| \geq k)$$

- same test set: $D_i = X_i - Y_i$

paired comparison

$p \simeq 0.6$

$t_{obs} = 0.7$



$X_1, \ldots, X_n$ i.i.d. $f(x; \theta)$; $\hat{\theta}(X_n)$ is maximum likelihood estimate.

$$(\hat{\theta} - \theta)/\widehat{\text{se}} \,\dot\sim\, N(0, 1)$$

$\hat{\theta} \,\dot\sim\, N(\theta, j^{-1}(\hat{\theta}))$

$j(\theta) = -\ell''(\hat{\theta})$

obs'd F. info

To test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ we could use

$$W = W(X_n) = (\hat{\theta} - \theta_0)/\widehat{\text{se}},$$

The rejection region will be $\{\boldsymbol{x} : |W(\boldsymbol{x})| > z_{\alpha/2}\}$, i.e. "reject" H_0 when $|W| \geq z_{\alpha/2}$

1.96 $\alpha = .05$

from $N(0,1)$

This test has approximate size α:

$$\Pr_{H_0}(|W| > z_{\alpha/2}) \doteq \alpha.$$

$\Pr_\theta\{|W| > z_{\alpha/2}\} = \text{power}(\theta)$

$$= 1 - P_{\theta}\left\{ -z_{\alpha/2} \le \frac{\hat\theta - \theta_0}{\hat{se}} \le z_{\alpha/2} \right\}$$

$$= 1 - P_{\theta}\left\{ \theta_0 - z_{\alpha/2}\,\hat{se} \le \hat\theta \le \theta_0 + z_{\alpha/2}\,\hat{se} \right\}$$

$$= 1 - P_{\theta}\left\{ \frac{\theta_0 - \theta - z_{\alpha/2}\hat{se}}{\hat{se}} \le \frac{\hat\theta - \theta}{\hat{se}} \le \underbrace{\frac{\theta_0 - \theta + z_{\alpha/2}\hat{se}}{\hat{se}}}_{} \right\}$$

$N(0,1)$

$$= 1 - P\left\{ \frac{\theta_0 - \theta}{\hat{se}} - z_{\alpha/2} \le Z \le \frac{\theta_0 - \theta}{\hat{se}} + z_{\alpha/2} \right\}$$

fn of θ for a given θ_0

AoS

152 10. Hypothesis Testing and p-values

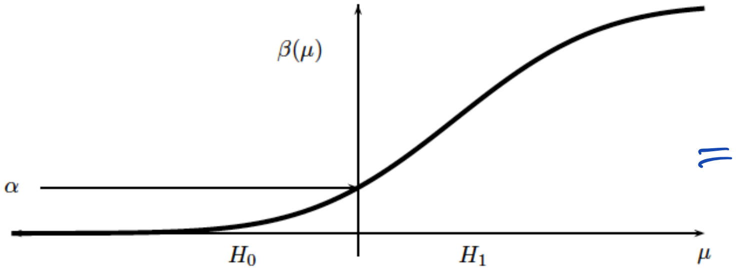


FIGURE 10.1. The power function for Example 10.2. The size of the test is the largest probability of rejecting H_0 when H_0 is true. This occurs at $\mu = 0$ hence the size is $\beta(0)$. We choose the critical value c so that $\beta(0) = \alpha$.

Let us consider the power of the Wald test when the null hypothesis is false.

10.6 Theorem. *Suppose the true value of θ is $\theta_\star \neq \theta_0$. The power $\beta(\theta_\star)$ — the probability of correctly rejecting the null hypothesis — is given (approximately) by*

$$1 - \Phi\left(\frac{\theta_0 - \theta_\star}{\hat{se}} + z_{\alpha/2}\right) + \Phi\left(\frac{\theta_0 - \theta_\star}{\hat{se}} - z_{\alpha/2}\right). \qquad (10.6)$$

$\sigma = (\mathrm{var}\, X_1)^{1/2}$

$se = \sigma/\sqrt{n}$ based $\hat\theta_n \sim N(\theta, j^{-1}(\theta))$

$\hat{se} = \sqrt{j^{-1}(\hat\theta)}$

- testing composite null hypothesis
- $X_1, \ldots, X_n$ i.i.d. $f(x; \boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$

$H_0: \theta \in \Theta_0$ "composite null"

dist- not completely spec'd

- composite $H_0 : \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}$ $r < p$
- notation $\boldsymbol{\theta} = (\psi, \boldsymbol{\lambda})$ (ϕ, τ)

parameter of interest

- $W_n =$ p.377

3 approximate test statistics $X_1, \ldots X_n$ iid $f(x; \underline{\theta})$

LR Stat
Wald
Score

- composite null hypothesis $H_0 : \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}$ $\qquad r < p$

- definition $\Lambda_n = \sup_{\theta \in \Theta} f(x_1, \ldots, x_n; \theta)$ $\qquad$ AoS Def 10.21 λ

$$\Lambda_n = \frac{\sup_{\theta \in \Theta} f(x_1, \ldots, x_n; \theta)}{\sup_{\theta \in \Theta_0} f(x_1, \ldots, x_n; \theta)} = \frac{L(\hat{\theta}; \underset{\sim}{x})}{L(\tilde{\theta}_0; \underset{\sim}{x})} \Leftarrow$$

$$1.)\ 2 \log \Lambda_n(\underset{\sim}{X}) \xrightarrow[H_0 : \theta \in \Theta_0]{d} \chi^2_r$$

$r = 1$

$$t(\underset{\sim}{X}) = 2 \log \Lambda_n(\underset{\sim}{X}) \qquad t^{crit} = \chi^2_{r, 1-\alpha}$$

Likelihood ratio statistic

- composite null hypothesis $H_0 : \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}$ $r < p$
- definition $\Lambda_n =$ AoS Def 10.21 λ
- Theorem MS 7.5, AoS 10.22

1. LR stat $2 \log \frac{L(\hat{\theta})}{L(\hat{\theta}_0)}$ ← full mle

mle under H_0

2. Wald : $(\hat{\psi} - \psi_0)^T j_p(\hat{\psi}) (\hat{\psi} - \psi_0) = W_n$

$\psi_0 = (\theta_{10}, \dots, \theta_{r0})$

$j_p(\psi) = -\ell_p''(\psi)$

$= -\frac{\partial^2}{\partial \psi^2} \ell(\psi, \hat{\lambda}_\psi)$

3. Score version

$\frac{\partial}{\partial x} \ell(\psi, \lambda; \underset{\sim}{x}) = 0 \Big|_{\lambda = \hat{\lambda}_\psi}$

Lagrange multiplier test

Rao test

$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \Lambda_n \geq \chi^2_{1-\alpha, r} \\ 0 & \text{o.w.} \end{cases}$

or $\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } W_n \geq \chi^2_{1-\alpha, r} \\ 0 & \text{o.w.} \end{cases}$

p-value is $Pr_{H_0}(\Lambda_n \geq \lambda_n^{obs})$ LRT

or $Pr_{H_0}(W_n \geq w_n^{obs})$ Wald test

… Example: logistic regression

$1, 0$

$P(y_i = 1) = \frac{e^{\beta_0}}{1+e^{\beta_0}}$

```
Boston.glmnull <- glm(crim2 ~ 1, family = binomial, data = Boston)

anova(Boston.glmnull, Boston.glm)
```

$\Pr(y_i = 1) = \frac{e^{\beta_0 + x_i^T \beta}}{1 + e^{\beta_0 + x_i^T \beta}}$

```
Analysis of Deviance Table

   Model 1: crim2 ~ 1
   Model 2: crim2 ~ (crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat + medv) - crim
  Resid. Df Resid. Dev Df Deviance
1       505     701.46
2       492     211.93 13   489.54
```

x_1 x_2

$H_0: \beta_1 = \cdots \beta_{13} = 0$

LR Test

$\Lambda_n = 489.$ $\quad r = 13$

p-value: $\Pr(\chi^2_{13} \geq 489)$

… Example: logistic regression

$H_0: \beta_2 = \beta_3 = \beta_5 = \beta_{12} = 0$

```
Boston.glmpart <- glm(crim2 ~ . - crim - indus - chas - rm - lstat,
                      data = Boston, family = binomial)

anova(Boston.glmpart, Boston.glm)
Analysis of Deviance Table
Model 1: crim2 ~ (crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat + medv) - crim - indus - chas -
    rm - lstat
Model 2: crim2 ~ (crim + zn + indus + chas + nox + rm + age + dis + rad +
    tax + ptratio + black + lstat + medv) - crim
  Resid. Df Resid. Dev Df Deviance
1       496     216.22
2       492     211.93  4   4.2891
```

13 x's

9 x's

x_2 x_3 x_5 x_{12}

$\Lambda_n = 4.3$ on 4 d.f

$\Pr(\chi^2_4 \geq 4.3) = .4$ (cutoff)

p-values

- The formal theory of testing imagines a decision to "reject H_0" or not, according as $X \in R$ or $X \notin R$, for some defined region $R \subset \mathcal{X}$ e.g. $|Z| > 1.96$
- This is useful for deriving the form of optimal tests, but not useful in practice.
- Doesn't distinguish between $Z = 1.97$ and $Z = 19.7$, for example.
- P-values give more precise information about the null hypothesis
- MS definition: $p(\boldsymbol{x}) = \inf\{\alpha : \phi_\alpha(\boldsymbol{x}) = 1\}$ 7.5
- AoS definition: p-value $= \inf\{\alpha : T(X_n) \in R_\alpha\}$ Def 10.11
- SM definition $p_{obs} = \Pr_{H_0}\{T(X_n) \geq t_{obs}\}$

Example: two-sample t-test

MS Ex.7.24

$X_1, \ldots X_m$ i.i.d. $N(\mu_1, \sigma^2)$, $Y_1, \ldots, Y_n$ i.i.d. $N(\mu_2, \sigma^2)$

$H_0 : \mu_1 = \mu_2$

LRT, Wald, score, exact

$p(t) = \mathrm{pr}_{H_0}(|T| > t)$

Example: Poisson

$X_1, \ldots X_n$ i.i.d. $Po(\lambda)$

$H_0 : \lambda = \lambda_0$

Example: logistic regression

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -34.103704   6.530014  -5.223 1.76e-07 ***
zn           -0.079918   0.033731  -2.369  0.01782 *
indus        -0.059389   0.043722  -1.358  0.17436
chas          0.785327   0.728930   1.077  0.28132
nox          48.523782   7.396497   6.560 5.37e-11 ***
rm           -0.425596   0.701104  -0.607  0.54383
age           0.022172   0.012221   1.814  0.06963 .
dis           0.691400   0.218308   3.167  0.00154 **
rad           0.656465   0.152452   4.306 1.66e-05 ***
tax          -0.006412   0.002689  -2.385  0.01709 *
ptratio       0.368716   0.122136   3.019  0.00254 **
black        -0.013524   0.006536  -2.069  0.03853 *
lstat         0.043862   0.048981   0.895  0.37052
medv          0.167130   0.066940   2.497  0.01254 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\hat\beta$ $\quad$ $\hat{se}$ $\quad$ $\hat\beta/\hat{se}$

$P_n(|N(0,1)| \geq z^{obs})$

$$\frac{\hat\beta_j - 0}{\hat{se}_j}$$ Wald st.

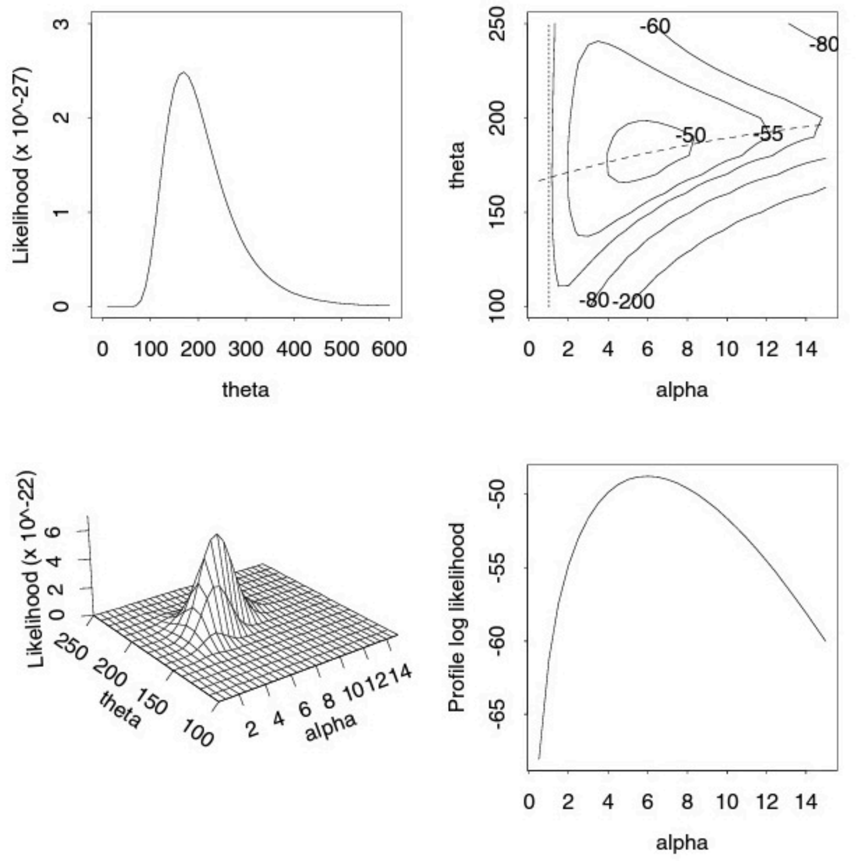
$\psi = \beta_j \quad H_0: \beta_j = 0$

$\lambda = \beta_{(j)}$

Likelihood ratio test revisited

AoS 10.6

4.1 · Likelihood 95



Sign test

- $X_1, \ldots, X_n$ i.i.d. $F(\cdot)$
- $H_0 : \mu = \mu_0$, $\mu = F^{-1}(1/2)$ median of distribution
- $H_1 : \mu > \mu_0$

both H composite

- test statistic

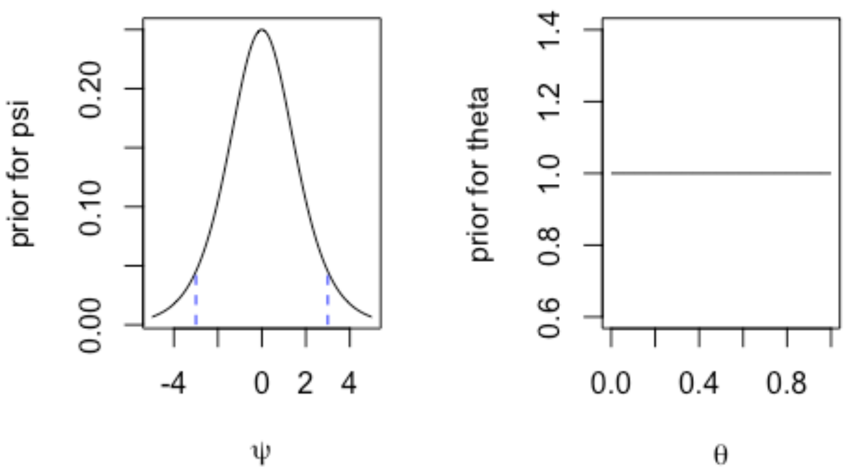
$$T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$$

- under H_0,

$$T \sim Binom(n, 1/2)$$

- p-value

$$p_{obs} = \mathrm{pr}_{H_0}(T \geq t_{obs}) = \sum_{r=t_{obs}}^{n} \binom{n}{r} \frac{1}{2^n} \doteq 1 - \Phi\left\{\frac{2(t_{obs} - n/2)}{n^{1/2}}\right\}.$$




```
> for(i in c(0,1,2,7))print(sum(gender==i))
[1] 831
[1] 1554
[1] 284
[1] 1331
> 1554+284
[1] 1838
> 1554/1838
[1] 0.8454842
> pbinom(1554, 1838, prob=1/2, lower = F)
[1] 1.19089e-212
> 1554+1331/2
[1] 2219.5
> 1838+1331
[1] 3169
> pbinom(2219, 3169, 1/2, lower=F)
[1] 3.634957e-116
```

```
> (.814*256-128)/sqrt(258*.814*.186)
[1] 12.86148
```

A Male Bias for Illusory Faces. To evaluate the robustness of the male bias observed in Exp. 1a, we plotted the distribution of all male and female ratings (i.e., excluding neutral responses) made by all participants as a function of image (Fig. 3*A*). Overall, there were significantly more male (81.4%) than female (18.6%) gender ratings [$z = 12.90$, $P = 2.24 \times 10^{-38}$, $n_{(images)}$ =

Power of the sign test

- $H_0 : F^{-1}(1/2) = \mu_0 \qquad H_1 : F^{-1}(1/2) > \mu_0$
- Test statistic $T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$
- $\mathrm{pr}_{H_0}(\text{reject } H_0) = \mathrm{pr}(T \geq c_\alpha \mid H_0) = \alpha \Rightarrow c_\alpha \approx n/2 - n^{1/2} z_\alpha/2$
- $\mathrm{pr}_{H_1}(\text{reject } H_0) = \mathrm{pr}(T \geq c_\alpha \mid H_1)$ Need distribution of T under H_1
- to calculate power we need values for μ and for F
- e.g. change to $H_1 : F^{-1}(1/2) = \mu_1$ $\mathrm{pr}_{F_{\mu_1}}(X > \mu_0)$
- SM assumes F is $N(\mu, \sigma^2)$, so $\delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$

$$\mathrm{pr}_{\mu_1}(T \geq c_\alpha) = \mathrm{pr}_{\mu_1}(T \geq n/2 - n^{1/2} z_\alpha/2) \doteq \Phi\left\{\frac{n\Phi(n^{-1/2}\delta) - n/2 + n^{1/2} z_\alpha}{[n\Phi(n^{-1/2}\delta)\{1 - \Phi(n^{-1/2}\}]}\right\}$$
$$\doteq \Phi\{z_\alpha + \delta(2/\pi)^{1/2}\}$$

- test based on $\bar{X}$ has power $\Phi(z_\alpha + \delta)$

… power of sign test

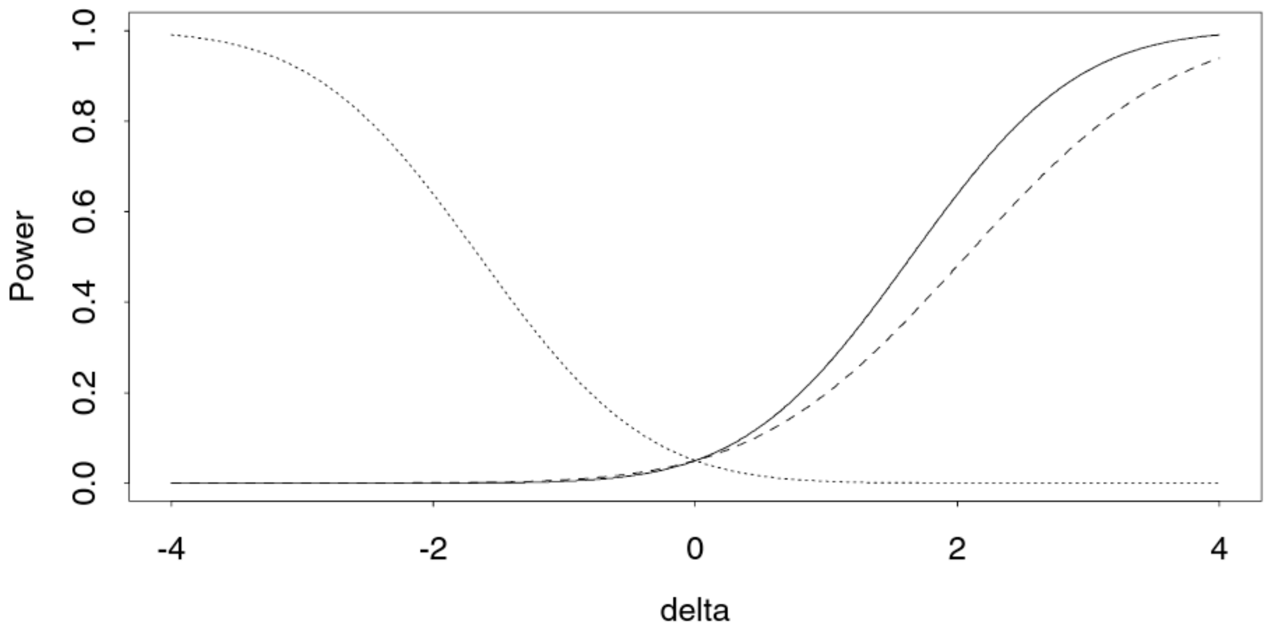


Figure 7.6 Power functions for a test of whether the mean of a $N(\mu, \sigma^2)$ random sample of size n equals μ_0 against the alternative $\mu = \mu_1$, as a function of $\delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$. The test size is $\alpha = 0.05$. The solid curve is the power function for a test of $\mu_1 > \mu_0$ based on $\overline{y}$, and the dashed line is the power function for the sign test. Both critical regions are of form $\overline{y} > t_\alpha$. The dotted curve is the power function for $\overline{y}$ when the critical region is $\overline{y} < t_\alpha$.

Other test statistics

$$X_1, \ldots, X_k \sim Mult(n; p_1, \ldots, p_k)$$

leukemia data (EH): $X_1, \ldots, X_{47}$; $Y_1, \ldots, Y_{25}$

AoS Ex. 10.20

```
oneline
          ALL     ALL.1     ALL.2     ALL.3     ALL.4     ALL.5     ALL.6     ALL.7
136 0.9186952 1.634002 0.4595867 0.6379664 0.3440379 0.8614784 0.5132176 0.9790902
        ALL.8     ALL.9    ALL.10    ALL.11   ALL.12    ALL.13    ALL.14    ALL.15   ALL.16
136 0.2105782 0.8016072 0.6006949 0.3614374 1.04632 0.9697635 0.4873159 0.4976364 1.101717
       ALL.17   ALL.18   ALL.19       AML     AML.1    AML.2    AML.3     AML.4     AML.5
136 0.8563937 0.661415 0.817711 0.7671718 0.9793741 1.425479 1.074389 0.9839282 0.9859271
        AML.6     AML.7   AML.8     AML.9   AML.10    AML.11    AML.12   AML.13    ALL.20
136 0.3247027 0.7110302 1.09625 0.9675151 0.975123 0.7775957 0.9472205 1.261352 0.5679544
       ALL.21    ALL.22    ALL.23    ALL.24    ALL.25    ALL.26   ALL.27    ALL.28
136 0.8462901 0.8838616 0.7239931 0.7327029 0.7823618 0.5435396 0.832537 0.5527333
       ALL.29    ALL.30    ALL.31    ALL.32   ALL.33    ALL.34    ALL.35    ALL.36
136 0.7327029 0.5510955 0.8214005 0.6418498 0.720798 0.5830999 0.7657568 0.5262976
      ALL.37    ALL.38    ALL.39   ALL.40    ALL.41    ALL.42    ALL.43    ALL.44
136 1.466999 0.5445589 0.5725049 1.362768 0.8533535 0.8132982 0.8538596 0.5689876
       ALL.45   ALL.46    AML.14    AML.15   AML.16   AML.17    AML.18    AML.19   AML.20
136 0.6930355 1.067526 0.9677959 0.9338141 1.138926 1.161753 0.6242354 0.6590103 1.215186
       AML.21   AML.22   AML.23    AML.24
136 0.9340861 1.310376 0.771426 0.7556606
```

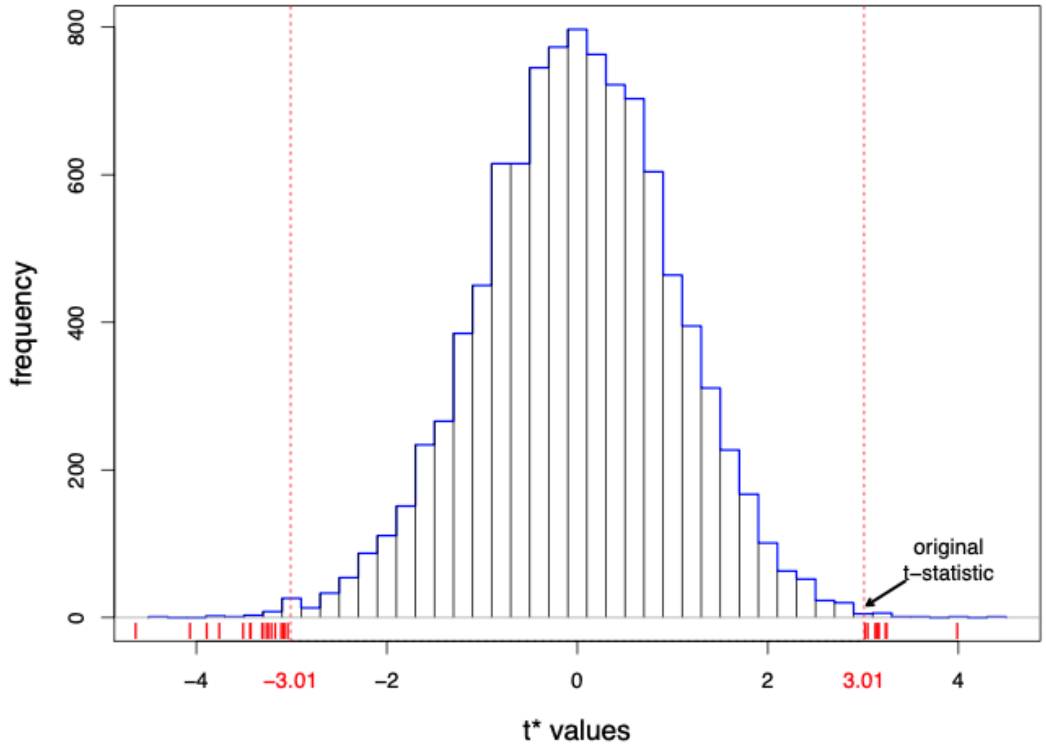


Figure 4.3 10,000 permutation t^*-values for testing **ALL** vs **AML**, for gene 136 in the **leukemia** data of Figure 1.3. Of these, 26 t^*-values (red ticks) exceeded in absolute value the observed t-statistic 3.01, giving permutation significance level 0.0026.

AoS 10.20 uses median

Choosing test statistics

1. Context

Choosing test statistics

1. Context

2. Optimal choice – Neyman-Pearson Lemma

Choosing test statistics

1. Context

2. Optimal choice – Neyman-Pearson Lemma

3. Pragmatic choice – likelihood-based statistics

Neyman-Pearson Lemma

- can we find the "best" test function $\phi(\boldsymbol{x})$ equivalently critical region R
- for testing H_0 vs H_1
- would like to minimize probability of two errors:

Neyman-Pearson Lemma

- can we find the "best" test function $\phi(\boldsymbol{x})$ — equivalently critical region R
- for testing H_0 vs H_1
- would like to minimize probability of two errors:

- fix $\mathrm{pr}(\text{reject } H_0 \mid H_0) \leq \alpha$, maximize $\mathrm{pr}(\text{reject } H_0 \mid H_1)$

Neyman-Pearson Lemma

- can we find the "best" test function $\phi(\boldsymbol{x})$ — equivalently critical region R
- for testing H_0 vs H_1
- would like to minimize probability of two errors:
- fix $\mathrm{pr}(\text{reject } H_0 \mid H_0) \leq \alpha$, maximize $\mathrm{pr}(\text{reject } H_0 \mid H_1)$
- Neyman-Pearson Lemma — MS Thm 7.2

$$\frac{f(\underline{x};\hat{\theta})}{f(\underline{x};\tilde{\theta}_0)} \quad \begin{matrix} \leftarrow \sup_{\Theta} \\ \leftarrow \sup_{\Theta_0} \end{matrix}$$

Neyman-Pearson Lemma

Suppose $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f(\boldsymbol{x})$. Under H_0, $f(\boldsymbol{X}) = f_0(\boldsymbol{x})$, and under H_1, $f(\boldsymbol{X}) = f_1(\boldsymbol{x})$.

simple simple

The test with test function

$$\frac{f(x;\theta_1)}{f(x;\theta_0)}$$

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } f_1(\boldsymbol{x}) > kf_0(\boldsymbol{x}), \\ 0 & \text{otherwise} \end{cases}$$

H_1 density

$$\frac{f_1(x)}{f_0(x)} > k$$

density H_0

(for some $0 < k < \infty$) is a most power test of H_0 vs H_1 at level

$$\alpha = \mathrm{E}_0\{\phi(\boldsymbol{X})\}.$$

$E_0\{\psi(X) \leq \alpha\}$

$$\int \psi(x)(f_1 - kf_0) \leq \int \phi(x)(f_1 - kf_0)$$

$\psi(x) \leq 1$?

if $\phi(x) = 1$ $\quad f_1 - kf_0 \geq 0$ $\quad \psi(x)(f_1 - kf_0) \leq 0$!

0 $\quad < 0$ $\quad$ +ve $\quad$ -ve $\quad$ yes bec

$$\int (\psi - \phi) f_1 \leq k \int (\psi - \phi) f_0$$

≥ 0 $\quad \leq 0$

$E_1\psi - E_1\phi \leq 0$

$E_1\psi \leq E_1\phi$