



HAL
open science

Statistics

Julien Lesouple

► **To cite this version:**

| Julien Lesouple. Statistics. Master. France. 2023. hal-04794782

HAL Id: hal-04794782

<https://enac.hal.science/hal-04794782v1>

Submitted on 21 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistics

ASNAT23

Julien Lesouple (TELECOM/SIGNAV)

julien.lesouple@enac.fr



Plan

① Estimation

② Statistical Hypothesis Testing

Plan

① Estimation

Statistical Modeling

Cramér-Rao bound

Maximum likelihood

Bayesian Estimation

Confidence Interval

② Statistical Hypothesis Testing

Plan

① Estimation

Statistical Modeling

Cramér-Rao bound

Maximum likelihood

Bayesian Estimation

Confidence Interval

② Statistical Hypothesis Testing

Statistical model

Observations

We have n real observations

$$x_1, \dots, x_n$$

Samples

We associate n **iid** random variables to the

$$X_1, \dots, X_n$$

The observations are realisations of independent random variables with the same underlying distribution

Statistical model

Goal: infer properties of the underlying distribution from the observations

Estimator

An estimator of a parameter θ is a function of the random variables associated to the observations

$$\hat{\theta}(X_1, \dots, X_n) \text{ or } \hat{\theta}_n \text{ or } \hat{\theta}$$

For instance, to estimate the mean of a distribution

$$\theta = E[X] \rightarrow \hat{\theta} = \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$$

How to compare different estimators? Is there a best choice?

Quality of an estimator

For $\theta \in \mathbb{R}$

Bias (systematic error)

The bias of an estimator is defined as

$$b_n(\theta) = E \left[\hat{\theta}_n \right] - \theta$$

Example: sample mean

Quality of an estimator

For $\theta \in \mathbb{R}$

Variance (convergence speed)

The variance of an estimator is defined as

$$v_n(\theta) = \text{Var}(\hat{\theta}_n) = E \left[\left(\hat{\theta}_n - E[\hat{\theta}_n] \right)^2 \right] = E[\hat{\theta}_n^2] - E[\hat{\theta}_n]^2$$

Example: sample mean

Quality of an estimator

For $\theta \in \mathbb{R}$

Mean square error (accuracy)

The mean square error (MSE) of an estimator is defined as

$$e_n(\theta) = E \left[\left(\hat{\theta}_n - \theta \right)^2 \right] = v_n(\theta) + b_n^2(\theta)$$

Example: sample mean

Convergence of an estimator

For $\theta \in \mathbb{R}$

Definition

An estimator $\hat{\theta}_n$ is said to be convergent if

$$b_n(\theta) \xrightarrow[n \rightarrow +\infty]{} 0$$

$$v_n(\theta) \xrightarrow[n \rightarrow +\infty]{} 0$$

Example: sample mean

Quality of an estimator

For $\theta \in \mathbb{R}^p$

Bias

The bias of an estimator is generalized as

$$b_n(\theta) = E \left[\hat{\theta}_n \right] - \theta \in \mathbb{R}^p$$

Covariance matrix

The variance of an estimator is generalized as

$$v_n(\theta) = E \left[\left(\hat{\theta}_n - E \left[\hat{\theta}_n \right] \right) \left(\hat{\theta}_n - E \left[\hat{\theta}_n \right] \right)^T \right] \in \mathbb{R}^{p \times p}$$

Exercise

We have $X_k \sim \mathcal{N}(m, \sigma^2)$ with known σ^2 . We want to estimate $\theta = m$. Compare the two estimators

- $\hat{\theta}_n = \frac{1}{n} \sum_{k=1}^n X_k$
- $\tilde{\theta}_n = \frac{2}{n(n+1)} \sum_{k=1}^n kX_k$

Solution

Bias for $\hat{\theta}_n$

$$\begin{aligned} b_{1,n}(m) &= E[\hat{\theta}_n] - m \\ &= E\left[\frac{1}{n} \sum_{k=1}^n X_k\right] - m \\ &= \left(\frac{1}{n} \sum_{k=1}^n E[X_k]\right) - m \\ &= \left(\frac{1}{n} \sum_{k=1}^n m\right) - m \\ &= 0 \end{aligned}$$

Solution

Bias for $\tilde{\theta}_n$

$$\begin{aligned} b_{2,n}(m) &= E \left[\tilde{\theta}_n \right] - m \\ &= E \left[\frac{2}{n(n+1)} \sum_{k=1}^n kX_k \right] - m \\ &= \left(\frac{2}{n(n+1)} \sum_{k=1}^n kE[X_k] \right) - m \\ &= \left(\frac{2}{n(n+1)} \sum_{k=1}^n km \right) - m \\ &= \left(\frac{2}{n(n+1)} \frac{n(n+1)}{2} \right) m - m \\ &= 0 \end{aligned}$$

Solution

Variance for $\hat{\theta}_n$

$$\begin{aligned}v_{1,n}(m) &= \text{Var}(\hat{\theta}_n) \\ &= \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) \\ &= \frac{1}{n^2} \text{Var}\left(\sum_{k=1}^n X_k\right) \\ &= \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) \quad (\text{independence}) \\ &= \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

Solution

Variance for $\tilde{\theta}_n$

$$\begin{aligned}v_{2,n}(m) &= \text{Var} \left(\tilde{\theta}_n \right) \\&= \text{Var} \left(\frac{2}{n(n+1)} \sum_{k=1}^n kX_k \right) \\&= \left(\frac{2}{n(n+1)} \right)^2 \text{Var} \left(\sum_{k=1}^n kX_k \right) \\&= \frac{4}{n^2(n+1)^2} \sum_{k=1}^n \text{Var} (kX_k) \\&= \frac{4}{n^2(n+1)^2} \sum_{k=1}^n k^2 \sigma^2\end{aligned}$$

Solution

Variance for $\tilde{\theta}_n$

$$\begin{aligned}v_{2,n}(m) &= \frac{4}{n^2(n+1)^2} \frac{n(n+1)(2n+1)}{6} \sigma^2 \\ &= \frac{2\sigma^2}{3} \frac{(2n+1)}{n(n+1)} \\ &= \mathcal{O}\left(\frac{4\sigma^2}{3n}\right)\end{aligned}$$

Solution

So $b_{1,n}(m) \xrightarrow[n \rightarrow +\infty]{} 0$ and $v_{1,n}(m) \xrightarrow[n \rightarrow +\infty]{} 0 \Rightarrow \hat{\theta}_n$ is convergent

$b_{2,n}(m) \xrightarrow[n \rightarrow +\infty]{} 0$ and $v_{2,n}(m) \xrightarrow[n \rightarrow +\infty]{} 0 \Rightarrow \tilde{\theta}_n$ is convergent

Moreover

$$\begin{aligned} \frac{v_{1,n}(m)}{v_{2,n}(m)} &= \frac{\sigma^2}{n} \frac{3}{2\sigma^2} \frac{n(n+1)}{(2n+1)} \\ &= \frac{3n+3}{4n+2} < 1 \text{ for } n > 1 \end{aligned}$$

So $v_{1,n}(m) < v_{2,n}(m)$ for $n > 1$

Likelihood

Remember the distribution of the samples is parametrized by θ

Definition

The likelihood of a parameter θ given the observations x_1, \dots, x_n is defined as

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \theta) &= \begin{cases} P_{X_1 \dots X_n}(x_1, \dots, x_n; \theta) & \text{if } X_k \text{ discrete} \\ f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta) & \text{if } X_k \text{ continuous} \end{cases} \\ &= \begin{cases} \prod_{k=1}^n P_X(x_k; \theta) & \text{if } X_k \text{ discrete} \\ \prod_{k=1}^n f_X(x_k; \theta) & \text{if } X_k \text{ continuous} \end{cases} \quad (\text{iid}) \end{aligned}$$

Sometimes: useful to work on the log-likelihood

$$\ln \mathcal{L}(x_1, \dots, x_n; \theta)$$

Plan

① Estimation

Statistical Modeling

Cramér-Rao bound

Maximum likelihood

Bayesian Estimation

Confidence Interval

② Statistical Hypothesis Testing

Cramér-Rao inequality

Bound on the variance of all the possible estimators of $\theta \in \mathbb{R}$

Theorem

If the log-likelihood is twice differentiable and if the support of the distribution does not depend on θ , then for any estimator $\hat{\theta}_n$ we have the inequality

$$\begin{aligned}\text{Var}(\hat{\theta}_n) &\geq \frac{[1 + b'_n(\theta)]^2}{-E\left[\frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta^2}\right]} = \text{CRB}(\theta) \\ &\geq \frac{[1 + b'_n(\theta)]^2}{-nE\left[\frac{\partial^2 \ln \mathcal{L}(X; \theta)}{\partial \theta^2}\right]} \quad (\text{iid})\end{aligned}$$

Counter-example: $\mathcal{U}([0, \theta])$

Efficient estimator

Definition

An estimator $\hat{\theta}_n$ is said to be efficient if it is unbiased and if

$$\text{Var}(\hat{\theta}_n) = \text{CRB}(\theta)$$

Property

If an efficient estimator exists, it is unique

Example: $X_k \sim \mathcal{N}(m, \sigma^2)$

Cramér-Rao inequality

Generalization to estimators of $\theta \in \mathbb{R}^q$

Theorem

For **unbiased** estimators $\hat{\theta}$ of θ , we have

$$\text{Cov}(\hat{\theta}) \geq I_n^{-1}(\theta)$$

where $A \geq B$ means $A - B$ is semi definite positive and

$$I_n(\theta)_{ij} = E \left[-\frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta_i \partial \theta_j} \right]$$

In particular

$$\text{Var}(\theta_i) \geq [I_n^{-1}(\theta)]_{ii}$$

Example

$$X_k \sim \mathcal{N}(m, \sigma^2) \text{ and } \boldsymbol{\theta} = (m, \sigma^2)^T$$

Plan

① Estimation

Statistical Modeling

Cramér-Rao bound

Maximum likelihood

Bayesian Estimation

Confidence Interval

② Statistical Hypothesis Testing

Maximum likelihood

Definition

The maximum likelihood estimator is defined as the value maximizing the likelihood with respect to θ

$$\hat{\theta}_{ML} = \arg \max_{\theta} \mathcal{L}(X_1, \dots, X_n; \theta)$$

Maximum likelihood

How to find the MLE?

If $\theta \in \mathbb{R}$ and the support of the distribution does not depend on θ and likelihood twice differentiable

$$\frac{\partial \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta} = 0 \text{ or } \frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta} = 0$$

then ensures it is a max

$$\frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta} \geq 0 \text{ or } \frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial^2 \theta} < 0$$

If $\theta \in \mathbb{R}^q$, for $i = 1, \dots, q$

$$\frac{\partial \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta_i} = 0 \text{ or } \frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta_i} = 0$$

Examples

Find the following MLE

- 1 $X_k \sim \mathcal{P}(\lambda)$ with $\theta = \lambda$
- 2 $X_k \sim \mathcal{N}(m, \sigma^2)$ with $\theta = (m, \sigma^2)^T$

Solution

In the first case we have $\theta > 0$

$$P_X(x_k; \theta) = e^{-\theta} \frac{\theta^{x_k}}{x_k!}, \quad x_k \in \mathbb{N}$$

$$\begin{aligned} \mathcal{L}(X_1, \dots, X_n; \theta) &= \prod_{k=1}^n P_X(X_k; \theta) \\ &= \prod_{k=1}^n e^{-\theta} \frac{\theta^{X_k}}{X_k!} \\ &= e^{-n\theta} \frac{\prod_{k=1}^n \theta^{X_k}}{\prod_{k=1}^n (X_k!)} \\ &= e^{-n\theta} \frac{\theta^{\sum_{k=1}^n X_k}}{\prod_{k=1}^n (X_k!)} \end{aligned}$$

Solution

Therefore

$$\begin{aligned}\ln \mathcal{L}(X_1, \dots, X_n; \theta) &= \ln \left(e^{-n\theta} \frac{\theta^{\sum_{k=1}^n X_k}}{\prod_{k=1}^n (X_k!)} \right) \\ &= -n\theta + \ln \left(\theta^{\sum_{k=1}^n X_k} \right) - \ln \left(\prod_{k=1}^n (X_k!) \right) \\ &= -n\theta + \sum_{k=1}^n X_k \ln(\theta) - \ln \left(\prod_{k=1}^n (X_k!) \right)\end{aligned}$$

The support of the distribution (\mathbb{R}^{+*}) does not depend on θ and $\ln \mathcal{L}(X_1, \dots, X_n; \theta)$ is clearly twice differentiable with respect to θ .

Solution

We have

$$\frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta} = -n + \frac{1}{\theta} \sum_{k=1}^n X_k$$

and

$$\frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta^2} = -\frac{1}{\theta^2} \sum_{k=1}^n X_k$$

therefore

$$\frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \hat{\theta}_{ML})}{\partial \theta} = 0 \Leftrightarrow \hat{\theta}_{ML} = \frac{1}{n} \sum_{k=1}^n X_k$$

Solution

And we check

$$\frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \theta} = -n + \frac{1}{\theta} \sum_{k=1}^n X_k$$

and

$$\begin{aligned} \frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \hat{\theta}_{ML})}{\partial \theta^2} &= -\frac{1}{\hat{\theta}_{ML}^2} \sum_{k=1}^n X_k \\ &= -\frac{n^2}{\sum_{k=1}^n X_k} < 0 \end{aligned}$$

Hence

$$\hat{\theta}_{ML} = \frac{1}{n} \sum_{k=1}^n X_k$$

Solution

In the second case, we have $\theta = (m, \sigma^2)^T \in \mathbb{R} \times \mathbb{R}^{*+}$ and

$$f_X(x_k; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_k - m)^2\right), \quad x_k \in \mathbb{R}$$

$$\begin{aligned}\mathcal{L}(X_1, \dots, X_n; \theta) &= \prod_{k=1}^n f_X(X_k; \theta) \\ &= \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_k - m)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^n (X_k - m)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{k=1}^n (X_k - m)^2\right)\end{aligned}$$

Solution

So

$$\ln \mathcal{L}(X_1, \dots, X_n; \theta) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^n (X_k - m)^2$$

For the mean we have

$$\begin{aligned} \frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial m} &= -\frac{1}{2\sigma^2} \sum_{k=1}^n 2(X_k - m)(-1) \\ &= \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - m) \end{aligned}$$

Solution

So if \hat{m}_{ML} is an extremum

$$\begin{aligned} \left. \frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial m} \right|_{m=\hat{m}_{ML}} = 0 &\Leftrightarrow \sum_{k=1}^n (X_k - \hat{m}_{ML}) = 0 \\ &\Leftrightarrow \sum_{k=1}^n X_k = n\hat{m}_{ML} \\ &\Leftrightarrow \frac{1}{n} \sum_{k=1}^n X_k = \hat{m}_{ML} \end{aligned}$$

and

$$\left. \frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial m^2} \right|_{m=\hat{m}_{ML}} = \frac{1}{\sigma^2} \sum_{k=1}^n (-1) = -\frac{n}{\sigma^2} < 0$$

Solution

For the variance we have

$$\begin{aligned} \frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \sigma^2} &= -\frac{n}{2} 2\pi \frac{1}{2\pi\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{k=1}^n (X_k - m)^2 \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{k=1}^n (X_k - m)^2 \end{aligned}$$

So for an extremum

$$\begin{aligned} \left. \frac{\partial \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial \sigma^2} \right|_{\sigma^2 = \hat{\sigma}_{ML}^2} = 0 &\Leftrightarrow \frac{1}{2(\hat{\sigma}_{ML}^2)^2} \sum_{k=1}^n (X_k - m)^2 = \frac{n}{2\hat{\sigma}_{ML}^2} \\ &\Leftrightarrow \frac{1}{n} \sum_{k=1}^n (X_k - m)^2 = \hat{\sigma}_{ML}^2 \end{aligned}$$

Solution

And

$$\begin{aligned}
 \frac{\partial^2 \ln \mathcal{L}(X_1, \dots, X_n; \theta)}{\partial(\sigma^2)^2} \Big|_{\sigma^2 = \hat{\sigma}_{ML}^2} &= \frac{n}{2(\hat{\sigma}_{ML}^2)^2} - 2 \frac{1}{2(\hat{\sigma}_{ML}^2)^3} \sum_{k=1}^n (X_k - m)^2 \\
 &= \frac{n}{2(\hat{\sigma}_{ML}^2)^2} - \frac{1}{(\hat{\sigma}_{ML}^2)^3} n \hat{\sigma}_{ML}^2 \\
 &= -\frac{n}{2(\hat{\sigma}_{ML}^2)^2} < 0
 \end{aligned}$$

so

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \hat{m}_{ML})^2$$

Maximum likelihood

Properties

- it is asymptotically unbiased
- it is convergent
- it is asymptotically efficient: $\frac{\text{Var}(\hat{\theta}_i)}{[I_n^{-1}(\theta)]_{ii}} \xrightarrow{n \rightarrow +\infty} 1$
- functional invariance: if $\mu = h(\theta)$ with h bijective, then $\hat{\mu}_{ML} = h(\hat{\theta})$

Generalization

All the previous definitions and results generalize when the observations X_1, \dots, X_n are vectors $\mathbf{X}_1, \dots, \mathbf{X}_n$

Plan

① Estimation

Statistical Modeling

Cramér-Rao bound

Maximum likelihood

Bayesian Estimation

Confidence Interval

② Statistical Hypothesis Testing

Brainstorming

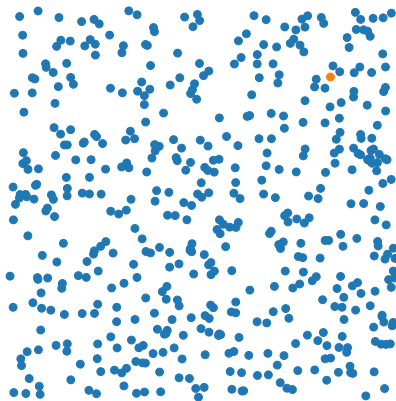
Medical test for a disease affecting 2‰ of the population

- Positive for 100% of ill people
- Negative for 95% of non ill people

You get tested positive. What is the probability you are indeed ill?

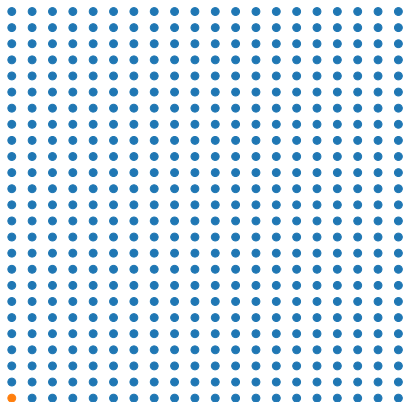
- 1 0-4.99%
- 2 5-49.99%
- 3 50-94.99%
- 4 95-100%

Brainstorming



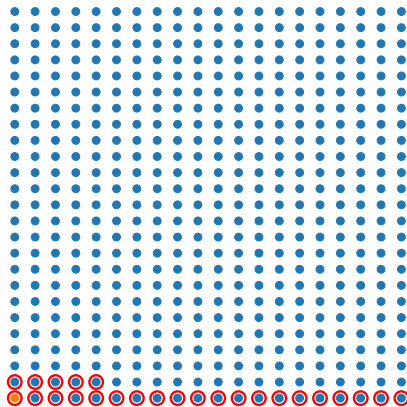
2‰: 500 people \rightarrow 1 sick

Brainstorming



2‰: 500 people \rightarrow 1 sick

Brainstorming



- 100% efficient for sick people: 1 tested positive
- 5% false alarm for healthy people: $(500-1) \times 0.05 \approx 24$ tested positive
- probability of being sick $\frac{1}{24+1} = 4\%$

Reminder

Framework:

- estimate parameters θ from n observations X_1, \dots, X_n
- we defined the likelihood $\mathcal{L}(X_1, \dots, X_n; \theta)$
 - how much a fixed θ explains the observations
- we defined the maximum likelihood estimator
 - the best θ explaining the observations

In the previous case

- $\theta \in \{0, 1\}$ (0: healthy, 1: sick)
- $X = 1$
- $\mathcal{L}(X; \theta) = P(X = 1|\theta) \begin{cases} P(X = 1|\theta = 1) = 1 \\ P(X = 1|\theta = 0) = 0.05 \end{cases}$
- we rather want to know $P(\theta = 1|X = 1)$

The Bayes theorem

Remember from probability class

Theorem

For two random variables \mathbf{X} and \mathbf{Y} , we have

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})f_{\mathbf{Y}}(\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}$$

continuous

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{P_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})P_{\mathbf{Y}}(\mathbf{y})}{P_{\mathbf{X}}(\mathbf{x})}$$

discrete

Brainstorming

In 1999, in England, a woman was judged because her two children suffered sudden death. The lawyer claimed: "Sudden death is extremely rare, so two times in a row is even more. I calculated 1 chance in 73 millions. So there is 1 chance in 73 millions that she is not guilty." She was put to jail. In your mind, was she guilty?

- 1 Yes
- 2 No

Exercise

We define $\theta \in \{0, 1\}$ (0: innocent, 1: guilty) and y the number of dead children. We want to find $P(\theta = 1|y = 2)$

$$P(\theta = 1|y = 2) = \frac{P(y = 2|\theta = 1)P(\theta = 1)}{P(y = 2)}$$

Exercise

We define $\theta \in \{0, 1\}$ (0: innocent, 1: guilty) and y the number of dead children. We want to find $P(\theta = 1|y = 2)$

$$P(\theta = 1|y = 2) = \frac{P(y = 2|\theta = 1)P(\theta = 1)}{P(y = 2)}$$

with

$$\begin{aligned}P(y = 2) &= P(y = 2|\theta = 1)P(\theta = 1) + P(y = 2|\theta = 0)P(\theta = 0) \\ &= P(y = 2|\theta = 1)P(\theta = 1) + P(y = 2|\theta = 0)(1 - P(\theta = 1))\end{aligned}$$

Exercise

We define $\theta \in \{0, 1\}$ (0: innocent, 1: guilty) and y the number of dead children. We want to find $P(\theta = 1|y = 2)$

$$P(\theta = 1|y = 2) = \frac{P(y = 2|\theta = 1)P(\theta = 1)}{P(y = 2)}$$

with

$$\begin{aligned} P(y = 2) &= P(y = 2|\theta = 1)P(\theta = 1) + P(y = 2|\theta = 0)P(\theta = 0) \\ &= P(y = 2|\theta = 1)P(\theta = 1) + P(y = 2|\theta = 0)(1 - P(\theta = 1)) \end{aligned}$$

and

$$P(y = 2|\theta = 1) = 1$$

$$P(y = 2|\theta = 0) = 1/73000000 \text{ (lawyer, but overestimated)}$$

$$P(\theta = 1) = 1/500000000 \text{ (UK statistics)}$$

Correction

$$P(\theta = 1|y = 2) = \frac{1 \times \frac{1}{500\,000\,000}}{1 \times \frac{1}{500\,000\,000} + \frac{1}{73\,000\,000} \times \left(1 - \frac{1}{500\,000\,000}\right)}$$

$$\approx 12.75 \%$$

Statisticians came to the court, she was freed after 3 years in jail
 This is why we are presumed innocent

A posteriori distribution

The parameter θ is now seen as a realization of a random variable Θ , with a prior distribution $f_{\Theta}(\theta)$.

Definition

Application of the Bayes theorem to the likelihood

$$f_{\Theta|\mathbf{X}_1 \dots \mathbf{X}_n}(\theta|\mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{f_{\mathbf{X}_1 \dots \mathbf{X}_n|\Theta}(\mathbf{x}_1, \dots, \mathbf{x}_n|\theta) f_{\Theta}(\theta)}{f_{\mathbf{X}_1 \dots \mathbf{X}_n}(\mathbf{x}_1, \dots, \mathbf{x}_n)}$$

\propto Likelihood \times prior

where \propto means “proportional to”

The Bayesian estimators

The Bayesian estimators minimize a cost $c(\theta, \hat{\theta})$ representing the error between θ and its estimate $\hat{\theta}$.

There are two main Bayesian estimators:

- the Minimum Mean Square Error (MMSE)
- the Maximum A Posteriori (MAP)

The MMSE estimator

Definition

Minimizes the Mean Square Error (MSE)

$$c(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \mathbb{E} \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right]$$

One can show

$$\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \mathbb{E}(\boldsymbol{\theta} | \mathbf{y}_1, \dots, \mathbf{y}_n)$$

It's the mean of the posterior distribution

The MAP estimator

Definition

Minimizes the cost $\mathbb{E} [c(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})]$ with

$$c(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \begin{cases} 1 & \text{if } \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\| > \Delta \\ 0 & \text{if } \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\| < \Delta \end{cases}$$

for Δ arbitrarily small. One can show

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} | \mathbf{y}_1, \dots, \mathbf{y}_n)$$

It's the value of $\boldsymbol{\theta}$ that maximizes de posterior distribution

Exercise

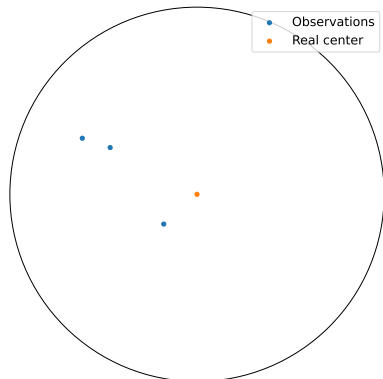
Suppose someone throws n arrows on a target with a marked center, and that you can only see the target from behind. Therefore, you can see the impact of the different arrows, but you don't know where the center is.

We model the position of the arrow as a Gaussian distribution, centred on the real center \mathbf{c} (the shooter aims perfectly, in mean), with a variance σ^2 (the precision) i.e.,

$$\mathbf{y}_i \sim \mathcal{N}(\mathbf{y}_i; \mathbf{c}, \sigma^2 I_2), \quad i = 1, \dots, n$$

- 1 what is the maximum likelihood estimator of the center?
- 2 do you think this estimator is appropriate when n is small?

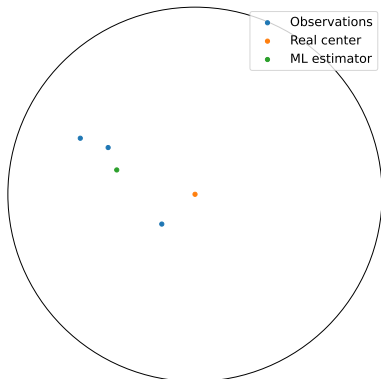
Exercise



ML for Gaussian mean: empirical mean

$$\hat{c} = \frac{1}{n} \sum_{i=1}^n y_i$$

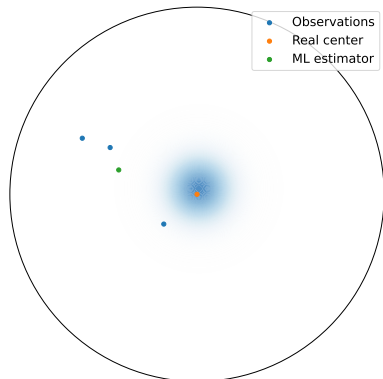
Exercise



ML for Gaussian mean: empirical mean

$$\hat{c} = \frac{1}{n} \sum_{i=1}^n y_i$$

Exercise



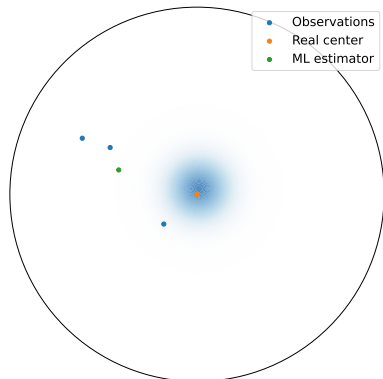
ML for Gaussian mean: empirical mean

$$\hat{\mathbf{c}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$$

We have an intuition on the center location, modelled as the prior

$$\mathbf{c} \sim \mathcal{N}(\mathbf{c}; \mathbf{c}_0, \sigma_0^2 \mathbf{I}_2)$$

Exercise



ML for Gaussian mean: empirical mean

$$\hat{\mathbf{c}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$$

We have an intuition on the center location, modelled as the prior

$$\mathbf{c} \sim \mathcal{N}(\mathbf{c}; \mathbf{c}_0, \sigma_0^2 I_2)$$

→ Derive the corresponding posterior and Bayesian estimators

Correction

We use Bayes rule to express the posterior as a function of \mathbf{c}

$$\begin{aligned}
 p(\mathbf{c}|\mathbf{y}_1, \dots, \mathbf{y}_n) &\propto \text{Likelihood} \times \text{prior} \\
 &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{c})^T (\mathbf{y}_i - \mathbf{c})\right) \exp\left(-\frac{1}{2\sigma_0^2} (\mathbf{c}_0 - \mathbf{c})^T (\mathbf{c}_0 - \mathbf{c})\right) \\
 &\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{y}_i^T \mathbf{y}_i - 2\mathbf{c}^T \mathbf{y}_i + \mathbf{c}^T \mathbf{c}) - \frac{1}{2\sigma_0^2} (\mathbf{c}_0^T \mathbf{c}_0 - 2\mathbf{c}_0^T \mathbf{c} + \mathbf{c}^T \mathbf{c})\right)
 \end{aligned}$$

I can remove multiplicative terms not depending in \mathbf{c}

$$p(\mathbf{c}|\mathbf{y}_1, \dots, \mathbf{y}_n) \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (-2\mathbf{c}^T \mathbf{y}_i + \mathbf{c}^T \mathbf{c}) - \frac{1}{2\sigma_0^2} (-2\mathbf{c}_0^T \mathbf{c} + \mathbf{c}^T \mathbf{c})\right)$$

Correction

I gather all the terms depending on \mathbf{c}

$$p(\mathbf{c} | \mathbf{y}_1, \dots, \mathbf{y}_n) \propto \exp \left(-\frac{1}{2} \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right) \mathbf{c}^T \mathbf{c} + \mathbf{c}^T \left(\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0 \right) \right)$$

I need to identify (if possible) this distribution with a known one. If we look at the multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\nu^2 I_2$

$$p(\mathbf{x}) \propto \exp \left(-\frac{1}{2\nu^2} \mathbf{x}^T \mathbf{x} + \frac{1}{\nu^2} \mathbf{x}^T \boldsymbol{\mu} \right)$$

which has the same shape as my posterior. I can thus identify

$$\frac{1}{\nu^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \quad \text{and} \quad \frac{1}{\nu^2} \boldsymbol{\mu} = \frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0$$

Correction

I can express the parameters

$$\nu^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right)^{-1} \quad \text{and} \quad \mu = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right)^{-1} \left(\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0 \right)$$

leading

$$\nu^2 = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \quad \text{and} \quad \mu = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \left(\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0 \right)$$

We conclude that the posterior distribution is

$$\mathbf{c} | \mathbf{y}_1, \dots, \mathbf{y}_n \sim \mathcal{N} \left(\mathbf{c}; \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \left(\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0 \right), \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} I_2 \right)$$

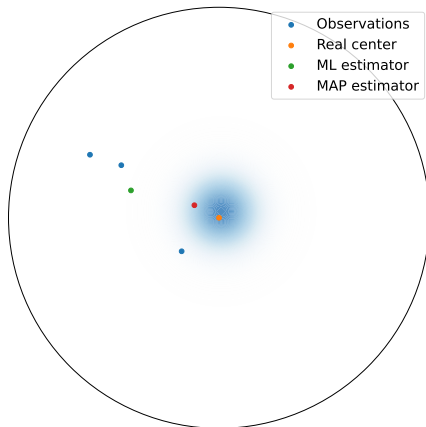
Correction

Note: we used a Gaussian prior for the mean of a Gaussian likelihood, and ended up with another Gaussian. We say that the Gaussian is a *conjugate* prior for the mean because the prior and the posterior are the same family.

We can deduce the Bayesian estimators. For the Gaussian distribution, the mean is also the value that maximizes the density. Therefore

$$\hat{\mathbf{c}}_{\text{MMSE}} = \hat{\mathbf{c}}_{\text{MAP}} = \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \left(\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0 \right)$$

Correction



Discussion

We have

$$\begin{aligned}
 \hat{\mathbf{c}}_{\text{MMSE}} &= \frac{\sigma^2 \sigma_0^2}{n\sigma_0^2 + \sigma^2} \left(\frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{y}_i + \frac{1}{2\sigma_0^2} \mathbf{c}_0 \right) \\
 &= \underbrace{\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}}_{\alpha_n} \underbrace{\frac{1}{n} \sum_{i=1}^n \mathbf{y}_i}_{\hat{\mathbf{c}}_{\text{ML}}} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \mathbf{c}_0 \\
 &= \alpha_n \hat{\mathbf{c}}_{\text{ML}} + (1 - \alpha_n) \mathbf{c}_0
 \end{aligned}$$

with $\alpha_n \rightarrow 1$ when $n \rightarrow +\infty$

The Bayesian estimator allow more importance to the prior when the number of observations n is small. On the contrary, when we have a large number of observations, the Bayesian estimator allows more importance to the maximum likelihood.

Conclusion

- How to introduce prior information into parametric estimation
- Bayes rule
- Posterior distribution (\propto likelihood \times prior)
- Bayesian estimators (MMSE and MAP)
- Application to the Gaussian mean

So far, we have seen how to give a point value for an estimator. But can we be more preservative and tell how much we are confident with this value?

Plan

① Estimation

Statistical Modeling

Cramér-Rao bound

Maximum likelihood

Bayesian Estimation

Confidence Interval

② Statistical Hypothesis Testing

Confidence interval

Definition

A confidence interval $[a, b]$ for $\theta \in \mathbb{R}$ is an interval such that

$$P[a < \theta < b] = \alpha$$

where a and b are functions of θ and $\alpha \in [0, 1]$ is the confidence level (usually $\alpha = 0.95$ or 0.99)

In practice, we find the distribution of the estimator to find such intervals

Exemples

- 1 $X_i \sim \mathcal{N}(m, \sigma^2)$, with m unknown and σ^2 is known
- 2 $X_i \sim \mathcal{N}(m, \sigma^2)$, with m unknown and σ^2 is unknown

Solution

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ is Gaussian with $\bar{X}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$ (property of Gaussian vectors)

Therefore

$$Z = \frac{\bar{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathcal{N}(0, 1)$$

Let $[a_\alpha, b_\alpha]$ a confidence interval with level α for the standard Gaussian distribution

$$P(a_\alpha < Z < b_\alpha) = \alpha$$

Solution

$$\begin{aligned}P(a_\alpha < Z < b_\alpha) &= P\left(a_\alpha < \frac{\bar{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} < b_\alpha\right) \\&= P\left(a_\alpha \sqrt{\frac{\sigma^2}{n}} < \bar{X}_n - m < b_\alpha \sqrt{\frac{\sigma^2}{n}}\right) \\&= P\left(-b_\alpha \sqrt{\frac{\sigma^2}{n}} < m - \bar{X}_n < -a_\alpha \sqrt{\frac{\sigma^2}{n}}\right) \\&= P\left(\bar{X}_n - b_\alpha \sqrt{\frac{\sigma^2}{n}} < m < \bar{X}_n - a_\alpha \sqrt{\frac{\sigma^2}{n}}\right)\end{aligned}$$

Solution

When the variance is unknown, we have seen $\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$ where $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ (property of Gaussian vectors) and \bar{X}_n and S_n^2 are independent. Therefore

$$U = \frac{Z}{\sqrt{\frac{S_n^2}{\sigma^2}}} \sim t_{n-1}$$

(Reminder: if $X \sim \mathcal{N}(0, 1)$ and $Y \sim \chi_n^2$ independent, $\frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$)

Let $[c_\alpha, d_\alpha]$ a confidence interval with level α for the Student distribution with $n - 1$ degrees of freedom

$$P(c_\alpha < U < d_\alpha) = \alpha$$

Solution

$$\begin{aligned}P(c_\alpha < U < d_\alpha) &= P\left(c_\alpha \sqrt{\frac{S_n^2}{\sigma^2}} < Z < d_\alpha \sqrt{\frac{S_n^2}{\sigma^2}}\right) \\&= P\left(c_\alpha \sqrt{\frac{S_n^2}{\sigma^2}} < \frac{\bar{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} < d_\alpha \sqrt{\frac{S_n^2}{\sigma^2}}\right) \\&= P\left(\bar{X}_n - d_\alpha \sqrt{\frac{S_n^2}{n}} < m < \bar{X}_n - c_\alpha \sqrt{\frac{S_n^2}{n}}\right)\end{aligned}$$

Remark

This Gaussian study case is very crucial because of the CLT!

Plan

① Estimation

② Statistical Hypothesis Testing

- Generalities

- ROC curves

- Neyman-Pearson

- Generalized Likelihood Ratio Test

Plan

① Estimation

② Statistical Hypothesis Testing

- Generalities

- ROC curves

- Neyman-Pearson

- Generalized Likelihood Ratio Test

Statistical hypothesis testing

Definition

A statistical test is a method of statistical inference to choose between several hypotheses H_0, H_1, \dots from observations x_1, \dots, x_n

In this class: only two hypotheses

For two hypotheses

A statistical test with two hypotheses consists in finding a test statistic $T(X_1, \dots, X_n)$ and a set Δ such that

H_0 is rejected if $T(X_1, \dots, X_N) \in \Delta$

H_0 is accepted if $T(X_1, \dots, X_N) \notin \Delta$

Statistical hypothesis testing

Definitions

- H_0 is the null hypothesis
- H_1 is the alternative hypothesis
- $\mathcal{R} = \{(x_1, \dots, x_n) | T(x_1, \dots, x_n) \in \Delta\}$ is the critical region
- an hypothesis is **simple** if we test a particular value of the parameter
- otherwise it is **composite**

Examples

- $H_0 : \theta = \theta_0$ is a simple hypothesis
- $H_0 : \theta > 0$ is a composite hypothesis

Statistical hypothesis testing

Definitions

- Type I error rate = probability of false alarm

$$\alpha = P_{FA} = P(\text{reject } H_0 | H_0 \text{ true})$$

- Type II error rate = probability of missed detection

$$\beta = P_{MD} = P(\text{reject } H_1 | H_1 \text{ true})$$

- power of the test = probability of detection $\pi = 1 - \beta$
- p -value $p(x_1, \dots, x_n) = \inf \{ \alpha \in (0, 1) | (x_1, \dots, x_n) \in \mathcal{R}_\alpha \}$

The p -value is the smallest value of α leading to the rejection of H_0

Example

$X_i \sim \mathcal{N}(m, \sigma^2)$, with known σ^2 . We want to test the two hypotheses

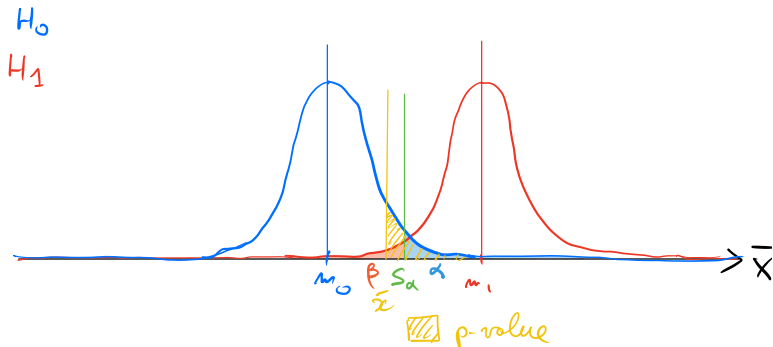
$$\begin{cases} H_0 : m = m_0 \\ H_1 : m = m_1 \end{cases}$$

where $m_1 > m_0$. For a fixed α

- 1 what statistic can we use to perform the test?
- 2 what is the critical region?
- 3 what is β ?
- 4 what is the power of the test?
- 5 what is the p -value of the test?

Solution

We want to test the value of the mean, hence the sample mean is a convenient statistic. Moreover, $\bar{X} \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$. We can compare the pdf of \bar{X} under the two hypotheses.



Solution

We can see that if the sample mean is too high, then the null hypothesis seems to be less likely. We propose the test

$$\text{reject } H_0 \text{ if } \bar{X} > S_\alpha$$

so

$$\mathcal{R}_\alpha = \left\{ (X_1, \dots, X_n) \in \mathbb{R}^n \left| \frac{1}{n} \sum_{i=1}^n X_i > S_\alpha \right. \right\}$$

Solution

To determine S_α , we write the definition of α

$$\begin{aligned}\alpha &= P(\text{reject } H_0 | H_0 \text{ true}) \\ &= P\left(\frac{1}{n} \sum_{i=1}^n X_i > S_\alpha \mid m = m_0\right) \\ &= P\left(\bar{X} > S_\alpha \mid \bar{X} \sim \mathcal{N}\left(m_0, \frac{\sigma^2}{n}\right)\right) \\ &= 1 - P\left(\bar{X} \leq S_\alpha \mid \bar{X} \sim \mathcal{N}\left(m_0, \frac{\sigma^2}{n}\right)\right) \\ &= 1 - P\left(\frac{\bar{X} - m_0}{\sqrt{\sigma^2/n}} \leq \frac{S_\alpha - m_0}{\sqrt{\sigma^2/n}} \mid \frac{\bar{X} - m_0}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)\right)\end{aligned}$$

Solution

Useful function: the inverse CDF of the standard Gaussian distribution

$\Phi^{-1}(x) =$ where

$$\Phi(u) = P(U \leq u | U \sim \mathcal{N}(0, 1))$$

is the CDF of the standard Gaussian distribution. We have

$$\begin{aligned} \alpha &= 1 - F_N \left(\frac{S_\alpha - m_0}{\sqrt{\sigma^2/n}} \right) \\ \Leftrightarrow F_N \left(\frac{S_\alpha - m_0}{\sqrt{\sigma^2/n}} \right) &= 1 - \alpha \\ \Leftrightarrow \frac{S_\alpha - m_0}{\sqrt{\sigma^2/n}} &= \Phi^{-1}(1 - \alpha) \Leftrightarrow S_\alpha = m_0 + \Phi^{-1}(1 - \alpha) \sqrt{\frac{\sigma^2}{n}} \end{aligned}$$

Solution

To find β , we write its definition

$$\begin{aligned}
 \beta &= P(\text{reject } H_1 | H_1 \text{ true}) \\
 &= P\left(\frac{1}{n} \sum_{i=1}^n X_i \leq S_\alpha \mid m = m_1\right) \\
 &= P\left(\frac{\frac{1}{n} \sum_{i=1}^n X_i - m_1}{\sqrt{\sigma^2/n}} \leq \frac{S_\alpha - m_1}{\sqrt{\sigma^2/n}} \mid \frac{\frac{1}{n} \sum_{i=1}^n X_i - m_1}{\sqrt{\sigma^2/n}} \sim \mathcal{N}(0, 1)\right) \\
 &= \Phi\left(\frac{S_\alpha - m_1}{\sqrt{\sigma^2/n}}\right) \\
 &= \Phi\left(\Phi^{-1}(1 - \alpha) + \frac{m_0 - m_1}{\sqrt{\sigma^2/n}}\right)
 \end{aligned}$$

Solution

To find the power, we write its definition

$$\begin{aligned}\pi &= 1 - \beta \\ &= 1 - \Phi \left(\Phi^{-1}(1 - \alpha) + \frac{m_0 - m_1}{\sqrt{\sigma^2/n}} \right)\end{aligned}$$

Same for the p -value

$$\begin{aligned}p &= \inf \{ \alpha \in (0, 1) \mid (x_1, \dots, x_n) \in \mathcal{R}_\alpha \} \\ &= \inf \left\{ \alpha \in (0, 1) \mid \frac{1}{n} \sum_{i=1}^n x_i > S_\alpha \right\} \\ &= \inf \left\{ \alpha \in (0, 1) \mid \frac{1}{n} \sum_{i=1}^n x_i > m_0 + \Phi^{-1}(1 - \alpha) \sqrt{\frac{\sigma^2}{n}} \right\}\end{aligned}$$

Solution

At the inf

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^n x_i &= m_0 + \Phi^{-1}(1-p) \sqrt{\frac{\sigma^2}{n}} \\ \Leftrightarrow \left(\frac{1}{n} \sum_{i=1}^n x_i - m_0 \right) \sqrt{\frac{n}{\sigma^2}} &= \Phi^{-1}(1-p) \\ \Leftrightarrow p &= 1 - \Phi \left(\left(\frac{1}{n} \sum_{i=1}^n x_i - m_0 \right) \sqrt{\frac{n}{\sigma^2}} \right)\end{aligned}$$

Plan

① Estimation

② Statistical Hypothesis Testing

- Generalities

- ROC curves

- Neyman-Pearson

- Generalized Likelihood Ratio Test

ROC curves

Definition

The Receiver Operator Characteristic (ROC) plots the power of the test versus the Type I error rate, i.e., the probability of detection versus the probability of false alarm)

$$\pi = f(\alpha)$$

In the previous example

$$\pi = 1 - \Phi \left(\Phi^{-1}(1 - \alpha) + (m_0 - m_1) \sqrt{\frac{n}{\sigma^2}} \right)$$

Plan

① Estimation

② Statistical Hypothesis Testing

Generalities

ROC curves

Neyman-Pearson

Generalized Likelihood Ratio Test

The Neyman-Pearson theorem

Framework

This theorem is defined for a test with simple hypotheses

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

Theorem: continuous random variables

For a fixed α , the test that minimizes β is

$$\text{Reject } H_0 \text{ if } \frac{\mathcal{L}(x_1, \dots, x_n; \theta_1)}{\mathcal{L}(x_1, \dots, x_n; \theta_0)} > S_\alpha$$

Example

$X_i \sim \mathcal{N}(m, \sigma^2)$, with known σ^2 . We want to test the two hypotheses

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m = m_1 \end{cases}$$

with $m_1 > m_0$. What is the most powerful test for a fixed α ?

Solution

We compute the likelihood

$$\begin{aligned}\mathcal{L}(x_1, \dots, x_n; m_1) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - m_1)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_1)^2\right)\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\mathcal{L}(x_1, \dots, x_n; m_1)}{\mathcal{L}(x_1, \dots, x_n; m_0)} &= \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_1)^2\right)}{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_0)^2\right)} \\ &= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_1)^2 + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_0)^2\right)\end{aligned}$$

Solution

$$\begin{aligned}
 \frac{\mathcal{L}(x_1, \dots, x_n; m_1)}{\mathcal{L}(x_1, \dots, x_n; m_0)} &= \exp \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - m_1)^2 - \sum_{i=1}^n (x_i - m_0)^2 \right) \right) \\
 &= \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n ((x_i - m_1)^2 - (x_i - m_0)^2) \right) \\
 &= \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (m_1^2 - m_0^2 - 2x_i(m_1 - m_0)) \right) \\
 &= \exp \left(-\frac{1}{2\sigma^2} \left(n(m_1^2 - m_0^2) - 2(m_1 - m_0) \sum_{i=1}^n x_i \right) \right)
 \end{aligned}$$

Solution

The Neyman-Pearson theorem leads

$$\text{Reject } H_0 \text{ if } \frac{\mathcal{L}(x_1, \dots, x_n; \theta_1)}{\mathcal{L}(x_1, \dots, x_n; \theta_0)} > S_\alpha$$

$$\exp\left(-\frac{1}{2\sigma^2}\left(n(m_1^2 - m_0^2) - 2(m_1 - m_0)\sum_{i=1}^n x_i\right)\right) > S_\alpha$$

$$-\frac{1}{2\sigma^2}\left(n(m_1^2 - m_0^2) - 2(m_1 - m_0)\sum_{i=1}^n x_i\right) > S'_\alpha$$

$$n(m_1^2 - m_0^2) - 2(m_1 - m_0)\sum_{i=1}^n x_i < S''_\alpha$$

$$-2(m_1 - m_0)\sum_{i=1}^n x_i < S'''_\alpha$$

Solution

We have $m_1 - m_0 > 0$ therefore the Neyman-Pearson test is formulated as

$$\text{Reject } H_0 \text{ if } \sum_{i=1}^n x_i > k_\alpha$$

Remark: earlier, we proposed

$$\text{reject } H_0 \text{ if } \bar{X} > S_\alpha$$

these are the same test as $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$

Neyman-Pearson recipe

To perform the Neyman-Pearson test for two **simple** hypothesis

- 1 find the test statistic with the likelihood ratio
- 2 find the critical region
- 3 determine the threshold S_α as a function of α
- 4 deduce β and π
- 5 numerical application to take a decision

The Neyman-Pearson theorem

Theorem: discrete random variables

For simple hypotheses and a fixed α , among all the tests with a Type I error rate $\leq \alpha$, the test that minimizes β is

$$\text{Reject } H_0 \text{ if } \frac{\mathcal{L}(x_1, \dots, x_n; \theta_1)}{\mathcal{L}(x_1, \dots, x_n; \theta_0)} > S_\alpha$$

When n is important, it is possible to resort to the central limit theorem

Example

$X_i \sim \mathcal{P}(\lambda)$. We want to test the two hypotheses

$$\begin{cases} H_0 : \lambda = \lambda_0 \\ H_1 : \lambda = \lambda_1 \end{cases}$$

with $\lambda_1 > \lambda_0$

Solution

We compute the likelihood

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \lambda_1) &= \prod_{i=1}^n P(X_i = x_i | \lambda_1) \\ &= \prod_{i=1}^n e^{-\lambda_1} \frac{\lambda_1^{x_i}}{x_i!} \\ &= e^{-n\lambda_1} \frac{\lambda_1^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \end{aligned}$$

Therefore

$$\frac{\mathcal{L}(x_1, \dots, x_n; m_1)}{\mathcal{L}(x_1, \dots, x_n; m_0)} = \frac{e^{-n\lambda_1} \frac{\lambda_1^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}}{e^{-n\lambda_0} \frac{\lambda_0^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}} = e^{-n(\lambda_1 - \lambda_0)} \left(\frac{\lambda_1}{\lambda_0} \right)^{\sum_{i=1}^n x_i}$$

Solution

The Neyman-Pearson theorem leads

$$\begin{aligned} \text{Reject } H_0 \text{ if } \frac{\mathcal{L}(x_1, \dots, x_n; \theta_1)}{\mathcal{L}(x_1, \dots, x_n; \theta_0)} &> S_\alpha \\ e^{-n(\lambda_1 - \lambda_0)} \left(\frac{\lambda_1}{\lambda_0} \right)^{\sum_{i=1}^n x_i} &> S_\alpha \\ \left(\frac{\lambda_1}{\lambda_0} \right)^{\sum_{i=1}^n x_i} &> S'_\alpha \\ \sum_{i=1}^n x_i \ln \left(\frac{\lambda_1}{\lambda_0} \right) &> S''_\alpha \end{aligned}$$

Solution

We have $\lambda_1 - \lambda_0 > 0$ therefore the Neyman-Pearson test is formulated as

$$\text{Reject } H_0 \text{ if } \sum_{i=1}^n x_i > k_\alpha$$

Plan

① Estimation

② Statistical Hypothesis Testing

- Generalities

- ROC curves

- Neyman-Pearson

- Generalized Likelihood Ratio Test

GLRT

Generalization of Neyman-Pearson for composite hypothesis (not necessarily the most powerful)

$$\begin{cases} H_0 : \boldsymbol{\theta} \in \Theta_0 \\ H_1 : \boldsymbol{\theta} \in \Theta_1 \end{cases}$$

Definition

$$\text{Reject } H_0 \text{ if } \frac{\mathcal{L}(x_1, \dots, x_n; \hat{\boldsymbol{\theta}}_1^{\text{ML}})}{\mathcal{L}(x_1, \dots, x_n; \hat{\boldsymbol{\theta}}_0^{\text{ML}})} > S_\alpha$$

where $\hat{\boldsymbol{\theta}}_0^{\text{ML}}$ (resp. $\hat{\boldsymbol{\theta}}_1^{\text{ML}}$) is the maximum likelihood estimator of $\boldsymbol{\theta}$ under H_0 (resp. H_1)

Remark: $\mathcal{L}(x_1, \dots, x_n; \hat{\boldsymbol{\theta}}_i^{\text{ML}}) = \sup_{\boldsymbol{\theta} \in \Theta_i} \mathcal{L}(x_1, \dots, x_n; \boldsymbol{\theta})$

Example

$X_i \sim \mathcal{N}(m, \sigma^2)$, with known σ^2 . We want to test the two hypotheses

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m \neq m_0 \end{cases}$$

What is the GLRT test?

Solution

$$\mathcal{L}(x_1, \dots, x_n; \hat{\theta}_i^{\text{ML}}) = \sup_{\theta \in \Theta_i} \mathcal{L}(x_1, \dots, x_n; \theta)$$

with $\Theta_0 = \{m_0\}$ and $\Theta_1 \neq \mathbb{R} \setminus \{m_0\}$. So

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \hat{\theta}_0^{\text{ML}}) &= \sup_{\theta \in \Theta_0} \mathcal{L}(x_1, \dots, x_n; \theta) \\ &= \mathcal{L}(x_1, \dots, x_n; m_0) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_0)^2 \right) \end{aligned}$$

Solution

And

$$\begin{aligned}\mathcal{L}(x_1, \dots, x_n; \hat{\theta}_1^{\text{ML}}) &= \sup_{\theta \in \Theta_1} \mathcal{L}(x_1, \dots, x_n; \theta) \\ &= \mathcal{L}(x_1, \dots, x_n; \bar{X}) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{X})^2 \right)\end{aligned}$$

Hence the GLRT

$$\text{Reject } H_0 \text{ if } \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{X})^2 + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_0)^2 \right) > S_\alpha$$

Solution

And

$$\begin{aligned} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{X})^2 + \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m_0)^2\right) &> S_\alpha \\ \Leftrightarrow -\sum_{i=1}^n (x_i^2 - 2x_i\bar{X} + \bar{X}^2) + \sum_{i=1}^n (x_i^2 - 2x_i m_0 + m_0^2) &> S'_\alpha \\ \Leftrightarrow 2n\bar{X}^2 - n\bar{X}^2 - 2n\bar{X}m_0 + nm_0^2 &> S''_\alpha \\ \Leftrightarrow n(\bar{X} - m_0)^2 &> k_\alpha \end{aligned}$$

with $\sqrt{n}(\bar{X} - m_0) \sim \mathcal{N}(0, 1) \rightarrow n(\bar{X} - m_0)^2 \sim \chi_1^2$