Seminar: Debating Statistical Inference Schools

Stefan Dietrich, Julian Rodemann & Thomas Augustin

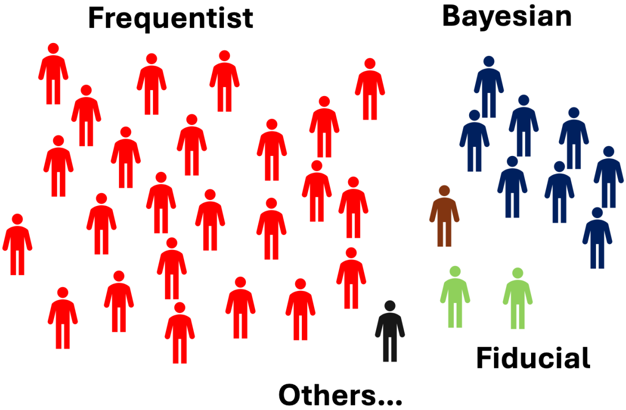
Department of Statistics, LMU Munich

Statistical Inference

"Statistical inference is the process of using data analysis to infer properties of an underlying distribution of probability."

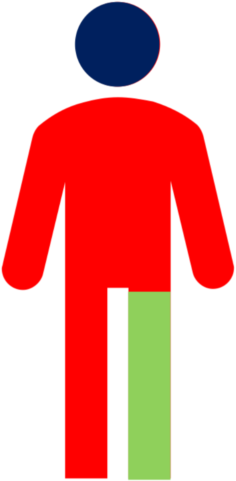
— *Oxford Dictionary of Statistics*

Schools



Schools

Not Mutually Exclusive



Probability

What Exactly is Probability?

- The coin will show heads with a 50% probability.
- There is a 30% probability of rain.
- The probability that 1860 Munich will win is 80%.
- ...

Frequentists

- Probability is Frequency
- Parameter is Fixed
- Parameter is Unknown [1]

[1] Comparative Statistical Inference Chapter 3.3

Frequentists

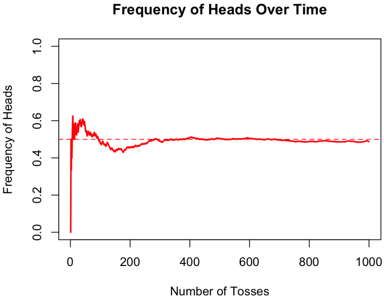
Frequency Equals Probability [2]

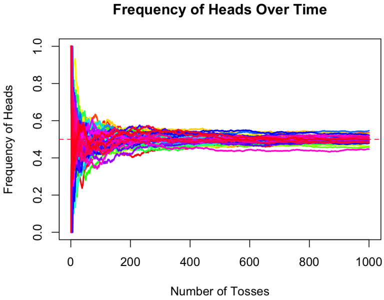
$$P(A) = \frac{\text{times A happens}}{\text{times Experiments}}$$

Calculate: Repeat experiment very often
Example: Coin toss

[2] Comparative Statistical Inference Chapter 3.3

Coin Toss Experiment





Coin Toss Experiment Results

- We repeat a coin toss a number of times and calculate the frequency of heads.
- **Expectation:** It gets to 0.5 over time.
- **Reality:**
  - Never exactly 0.5, often very close, sometimes off.
  - Gets better over time.

Law of Large Numbers

- The Law of Large Numbers states that as the number of trials increases, the sample mean will converge to the expected value. [3]
- In the context of a coin toss:
  - The expected value of heads is 0.5.
  - As we increase the number of tosses, the proportion of heads will approach 0.5.

Formula:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu \quad \text{as } n \to \infty$$

[3] Comparative Statistical Inference Chapter 5

Limitations of Infinite Repetitions

- **One-time Events:**
  - Some events occur only once (e.g., a unique historical event).
  - Cannot apply frequentist methods that assume repeated trials.
- **Non-Independent Trials:**
  - Trials may be dependent (e.g., in a time series).
  - Assumption of independence is violated.
- **Limited Resources:**
  - In practice, repeating an experiment infinitely is not feasible.
  - Resource constraints limit the number of trials. [4]

[4] TODO

Running example: blood pressure

We are analyzing systolic blood pressure and have data from 10 patients:

$$120, 125, 130, 110, 115, 140, 135, 128, 118, 122$$

The goals are to:

- Estimate the mean μ using different inference methods
- Construct intervals (Confidence, Credibility, Fiducial)
- Perform hypothesis tests: Is the mean greater than 125?

Frequentist Inference - Point Estimator

Point Estimator: The sample mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{120 + 125 + \cdots + 122}{10} = 124.3$$

This serves as our estimate for the population mean μ.

Frequentist Inference - Confidence Interval

Constructing the 95% Confidence Interval:

$$\bar{x} \pm z \cdot \frac{s}{\sqrt{n}} = 124.3 \pm 1.96 \cdot \frac{8.57}{\sqrt{10}}$$

$$\text{Confidence Interval: } [117.98, 130.62]$$

This interval means that if we repeated this process many times, 95% of such intervals would contain the true mean.[5]

[5] Comparative Statistical Inference Chapter 2

Frequentist Inference - Hypothesis Test

Hypothesis Test:

- Null Hypothesis: $H_0 : \mu \leq 125$
- Alternative Hypothesis: $H_A : \mu > 125$

$$z = \frac{\bar{x} - 125}{\frac{s}{\sqrt{n}}} = \frac{124.3 - 125}{\frac{8.57}{\sqrt{10}}} = -0.26$$

Since $z < z_{1-\alpha}$, we do not reject H_0.

Bayesian

- Based on Bayes' Theorem
- Parameter is a random variable
- Parameter is function[6]

[6] Comparative Statistical Inference Chapter 6

Theorem of Bayes

$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{P(X)}$$

- $P(\theta|X)$: Posterior probability of θ given Data X.
- $P(X|\theta)$: Likelihood of Data X given that θ is true.
- $P(\theta)$: Prior probability of θ.
- $P(X)$: Marginal likelihood of the Data X.[7]

[7] Comparative Statistical Inference Chapter 6

Posterior Probability $P(\theta|X)$

- Represents the updated probability of the hypothesis θ after observing evidence X. This is a function and not a number as in frequentism.
- Meaning: It reflects our belief about θ considering both prior information and new evidence.
- Challenges:
  - May be sensitive to the prior choice, leading to biases.
  - Computationally intensive to derive, especially with complex models.[8]

[8] Comparative Statistical Inference Chapter 6

Likelihood Function $P(X|\theta)$

- The probability of observing the data X given that the hypothesis θ is true.
- Meaning: Indicates how well the model with hypothesis θ explains the observed data.
- Challenges:
  - Requires an accurate model of the data generation process.[9]

[9] Comparative Statistical Inference Chapter 6

Prior Probability $P(\theta)$

- Represents the initial belief about the hypothesis θ before observing any evidence.
- Meaning: Captures existing knowledge or assumptions about θ and influences the posterior.
- Challenges:
  - Choice of prior can significantly impact the results.
  - Eliciting a subjective prior can be difficult and controversial.[10]

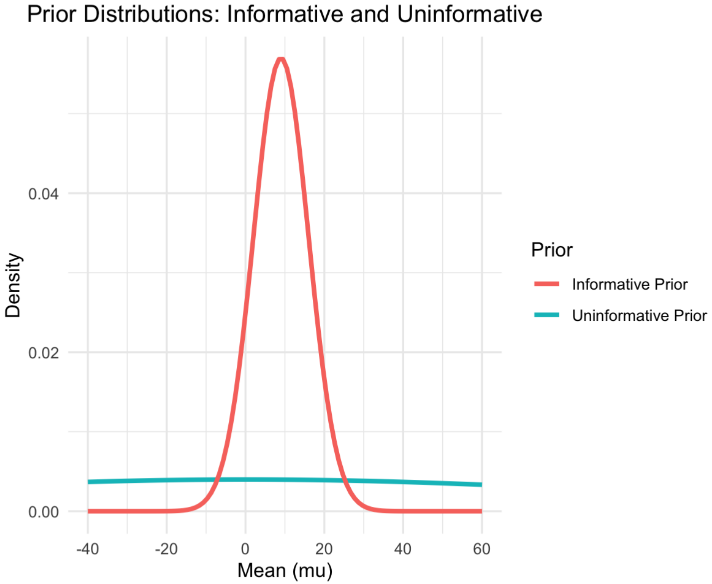
[10] Comparative Statistical Inference Chapter 6

Choosing the Prior

- Types of Priors:
  - **Informative Priors**:
    - Based on previous studies or expert knowledge.
    - Provides strong influence on the posterior.
  - **Uninformative (or Weak) Priors**:
    - Reflects minimal prior knowledge (e.g., uniform distribution).
    - Aims to let the data dominate the inference.[11]
- Importance of Reporting Priors:
  - In scientific work, it is essential to specify the chosen prior.
  - This transparency allows for discussion and scrutiny of the assumptions made during analysis.

[11] Comparative Statistical Inference Chapter 6

Informative vs. Uninformative Prior



Conjugate Priors

- Conjugate priors are a class of prior distributions that, when used in Bayesian analysis, yield a posterior distribution that is in the same family as the prior.
- This property simplifies the computation of the posterior distribution. [12]

[12] Comparative Statistical Inference Chapter 6

Examples of Conjugate Priors

- **Bernoulli Likelihood:** - Prior: Beta Distribution $\text{Beta}(\alpha, \beta)$ - Posterior: $\text{Beta}(\alpha + k, \beta + n - k)$ where k is the number of successes and n is the number of trials.
- **Normal Likelihood:** - Prior: Normal Distribution $N(\mu_0, \sigma_0^2)$ - Posterior: $N(\mu_n, \sigma_n^2)$[13]

[13] Comparative Statistical Inference Chapter 6

Updating the Mean for a Normal Distribution

Setup:

- Prior: $\theta \sim N(\mu_0, \sigma_0^2)$
- Likelihood: $x|\theta \sim N(\theta, \sigma^2)$

Posterior Distribution:

$$\theta|x \sim N(\mu_n, \sigma_n^2)$$

Updated Mean:

$$\mu_n = \frac{\frac{x}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

Updated Variance:

$$\sigma_n^2 = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

Interpretation:

- The posterior mean μ_n is a weighted average of the prior mean μ_0 and the observed data x, where the weights depend on their respective precisions. [14]

[14] Comparative Statistical Inference Chapter 6

Advantages of Conjugate Priors

- Simplifies calculations, making Bayesian analysis more tractable.
- Provides a clear understanding of how prior beliefs combine with data.

Marginal likelihood $P(X)$

- The total probability of observing the data X across all possible hypotheses θ.
- Meaning: Acts as a normalizing constant to ensure the posterior probability sums to one.
- Challenges:
  - Often difficult to compute in practice, especially in high-dimensional spaces. May require complex integration or approximation techniques.[15]

[15] Comparative Statistical Inference Chapter 6

Example: Marginal Likelihood

- The marginal likelihood with logistic regression and normal prior is:

$$P(X) = \int_{\Theta} \left(\prod_{i=1}^{n} \left(\frac{1}{1 + \exp(-x_i^\top \theta)} \right)^{y_i} \left(1 - \frac{1}{1 + \exp(-x_i^\top \theta)} \right)^{1-y_i} \right)$$

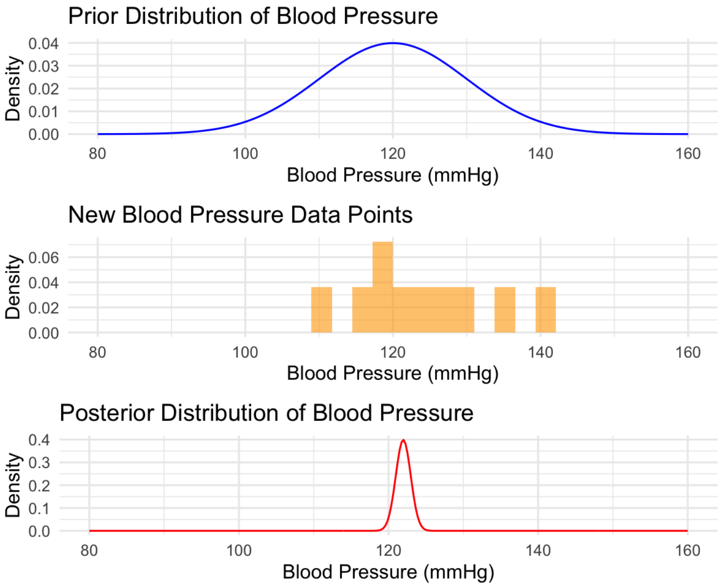
$$\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\theta - \mu)^2}{2\sigma^2} \right) d\theta$$

Approximation Methods for Marginal Likelihood

- Approximation Techniques:
  - **Monte Carlo Methods (MCMC)**:
    - Random sampling from the distribution to estimate the integral. [16]
  - **Variational Inference**:
    - Approximates the integral with optimization techniques to find simpler distributions.[17]
  - **Laplace Approximation**:
    - Uses a quadratic approximation around the mode of the posterior to estimate the integral.[18]

[16] Comparative Statistical Inference Chapter 6
[17] A Tutorial on Variational Bayes
[18] The Classical Laplace Method

Updating a Prior



Bayesian Inference - Point Estimator

Point Estimator: Posterior Mean $\mathbb{E}[\mu|X]$:

$$\mu_{\text{posterior}} = \frac{\frac{1}{\sigma^2}\bar{x} + \frac{1}{\tau^2}\mu_0}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}$$

Assuming $\mu_0 = 120$, $\tau^2 = 15^2$, $\bar{x} = 124.3$, and $\sigma^2 = 10^2$:

$$\mu_{\text{posterior}} = \frac{\frac{1}{100} \cdot 124.3 + \frac{1}{225} \cdot 120}{\frac{1}{100} + \frac{1}{225}} \approx 123.4$$

Bayesian Inference - Credibility Interval

Constructing the 95% Credibility Interval: This interval is derived from the posterior distribution:

$$[118.9, 128.0]$$

This means that there is a 95% probability that the true mean lies within this interval based on our prior belief and the data.[19]

[19] Comparative Statistical Inference Chapter 6

Bayesian Inference - Hypothesis Test

Hypothesis Test: We compute the Bayes factor:

$$\frac{P(\mu > 125 \mid \mathrm{X})P(\mu > 125)}{P(\mu \leq 125 \mid \mathrm{X})P(\mu \leq 125)} = 0.36$$

Since Bayes factor is less than 1, we do not reject H_0.

Comparing the Results

Frequentist

- What is μ?

  $\hat{\mu} = 124.3$ (Sample Proportion)

- Interval for μ?

  $(117.98, 130.62)$ (95% Confidence Int.)

- Conclusion

  Not rejected the Null

Bayesian

- What is μ?

  $\mathbb{E}(\mu|y) = 123.4$ (Posterior Mean)

- Interval for μ?

  $(118.9, 128.0)$ (95% Credible Int.)

- Conclusion

  Not rejected the Null

Inverse Probability

What is Inverse Probability?

- Inverse probability is a method of reasoning backward from observed data to infer the probability of underlying causes or parameters.
- It contrasts with direct probability, where we calculate the probability of data given certain parameters.
- Central to Bayesian statistics.[20]

[20] Comparative Statistical Inference Chapter 6

Fiducial Inference

- Background:
  - The term "fiducial"comes from the Latin for faith.
  - Ronald A. Fisher introduced fiducial inference in the 1930s as a way to derive statistical inferences without relying solely on traditional frequentist or Bayesian approaches.
- Fisher's Key Argument:
  - after estimating parameters from the data, we can construct a fiducial distribution for the parameters based on the observed data and the sampling distribution.
  - inferences about parameters without needing prior distributions
  - Doubts exist about the coherence of fiducial inference as a system of statistical inference [21]
  - Fiducial distributions lack additivity, preventing them from forming a valid probability (Lindeley) measure.[22]

[21] Comparative Statistical Inference Chapter 8.1

[22] Lindeley, D.V. (1958) Fiducial Distributions and Bayes' Theorem, Journal of the Royal Statistical Society

Fiducial vs Bayesian Inference

- Fiducial inference does not rely on priors, unlike Bayesian inference.
- Bayesian inference incorporates prior beliefs into the analysis.
- Fiducial argument aims for objectivity, avoiding subjective priors. [23]

[23] Comparative Statistical Inference Chapter 8.1

Fiducial Distribution

A random sample of size n from $N(\mu, \sigma_0^2)$ with known variance σ_0^2 has a sample mean $\bar{x}$, $\bar{X}$ is sufficient for μ and has distribution function $\Phi\left(\frac{(\bar{x}-\mu)\sqrt{n}}{\sigma_0}\right)$. Thus, the fiducial distribution of μ, for given $\bar{x}$, has density

$$g(\mu; x) = \sqrt{\left(\frac{n}{2\pi\sigma_0^2}\right)} \exp\left\{-\frac{n}{2\sigma_0^2}(\mu - \bar{X})^2\right\}$$

The fiducial interval and central confidence interval for μ are identical.[24]

[24] Comparative Statistical Inference Chapter 8.1

Interpretation of the Likelihood Function

The fiducial distribution can also be derived by attributing a probability interpretation to the likelihood function $p_\mu(x)$, normalized as:

$$p_\mu(x) \propto g(\mu; x)$$

If we interpret the likelihood function as measuring relative densities of μ, we arrive at the same probability distribution.
However, this re-interpretation does not always suffice.[25]

[25] Comparative Statistical Inference Chapter 8.1

Fiducial Inference - Point Estimator

Point Estimator: The sample mean (same as Frequentist):

$$\bar{x} = 124.3$$

This serves as the fiducial estimate for the population mean.

Fiducial Inference - Fiducial Interval

Fiducial Interval: This interval can be calculated similarly to the confidence interval:

$$[118.9, 128.0]$$

This interval provides a fiducial estimate for the true mean based on the observed data.

References

- Oxford Dictionary of Statistics (2024). Definition of Statistical Inference.
- Vic Barnett (1999), Comparative Statistical Inference (Third Edition)
- Lindeley, D.V. (1958) Fiducial Distributions and Bayes' Theorem, Journal of the Royal Statistical Society