MATH 5010 –Foundations of Statistical Theory and Probability

❖ Interval Estimation 2

-Evaluating Interval Estimators

Instructor: He Wang
Department of Mathematics
Northeastern University

1. Size and Coverage Probability

2. Test-Related Optimality

3. Bayesian Optimality

4. Loss Function Optimality

1. Size (length) and Coverage Probability

The **coverage probability** of a confidence interval (CI) is the probability that the interval contains the true parameter value.

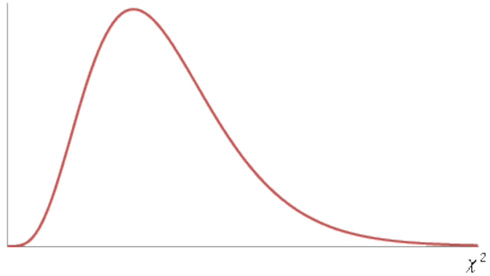
$$P_\theta\big(\theta \in C(X)\big) \approx 1 - \alpha$$

The "**size**" usually refers to the **length** (or expected length) of the interval:

Length: $L(X) = U(X) - L(X)$

Expected length: $E[U(X) - L(X)]$

Smaller size is Better: Among intervals with the same coverage, the shorter one is preferred (gives more precise estimation).



Example

Confidence interval for μ when $X_1, \ldots, X_n \sim Normal(\mu, \sigma^2)$ with σ known.

From Pivot construction:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim Normal(0,1)$$

For any constants a and b,

$$P(a \leq Z \leq b) = 1 - \alpha$$

gives a valid $(1 - \alpha)$ confidence interval for μ

$$\left\{\mu: \bar{X} - b\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - b\frac{\sigma}{\sqrt{n}}\right\}$$

The length is $Length = (b - a)\frac{\sigma}{\sqrt{n}}$

Fix α, which choice of a, b

Example (shortest interval**)**

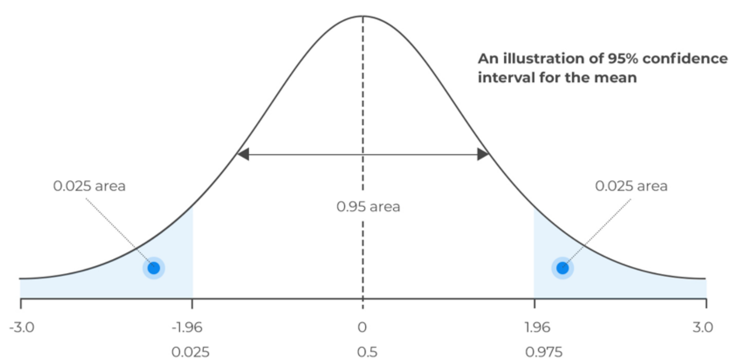
- Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, with σ^2 known.
- A **95% confidence interval for** μ is:

$$\bar{X} \pm z_{0.975} \frac{\sigma}{\sqrt{n}}$$

- Here:
  - **Coverage Probability**: Exactly 0.95 for all μ.
  - **Size (Length)**:

$$2 z_{0.975} \frac{\sigma}{\sqrt{n}}$$

which decreases as sample size n increases → more precision.



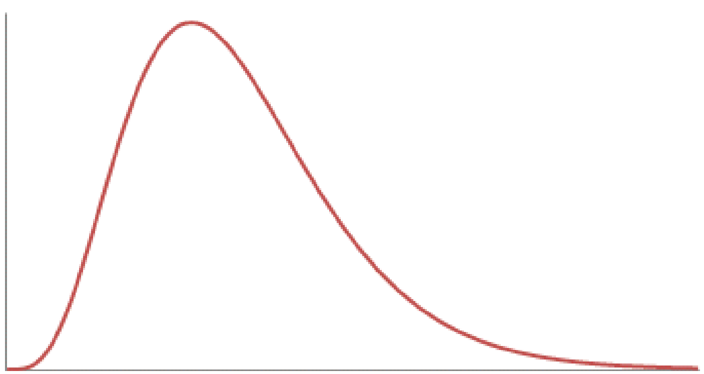
Shortest Interval with a Unimodal PDF

Theorem

Let $f(x)$ be a unimodal PDF. (i.e., there exist mode x^* such that $f(x)$ is non-decreasing for $x \leq x^*$, and $f(x)$ is non-decreasing for $x \geq x^*$.

If the interval $[a, b]$ satisfies:

- $\int_a^b f(x)\,dx = 1 - \alpha$ (coverage condition)
- $f(a) = f(b) > 0$ (equal boundary density condition)
- $a \leq x^* \leq b$, where x^* is the mode of $f(x)$ (interval contains the mode)



2. Test-Related Optimality

Constructing a $(1-\alpha)$ confidence interval can be seen as inverting the acceptance regions of level-α hypothesis tests.

Optimal tests translate into properties of *optimal confidence intervals*.

False coverage probability:

$$P_\theta\big(\theta' \in C(X)\big), \text{ for } \theta' \neq \theta$$

A $(1-\alpha)$ confidence set is said to be **Uniformly Most Accurate** (UMA) if it minimizes the probability of false coverage among all sets with the same coverage probability.

Theorem:

Let $X \sim f(x|\theta), \ \theta \in \mathbb{R}$

Let $A^*(\theta_0)$ be the UMP level-α acceptance region of a test of

$$H_0: \theta = \theta_0 \text{ v.s. } H_1: \theta > \theta_0$$

Let $C^*(X)$ be the $(1-\alpha)$ confidence set obtained by inverting these acceptance regions.

Then, for any other $(1-\alpha)$ confidence set C,

$$P_\theta\big(\theta' \in C^*(X)\big) \leq P_\theta\big(\theta' \in C(X)\big), \text{ for } \theta' < \theta$$

Remark: The confidence set obtained from the UMP test minimizes false coverage probability among all valid confidence sets → it is UMA.

Example: $X_1, \dots, X_n \sim Normal(\mu, \sigma^2)$ with σ known.

- UMA lower confidence bound inverted from the UMP test: $H_0: \theta = \theta_0$ v.s. $H_1: \theta > \theta_0$

$$C(x) = \left\{\mu: \mu \geq \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}\right\}$$

- Two-sided interval

$$C(x) = \left\{\mu: \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right\}$$

This is not UMA, because no UMP test exists for the two-sided hypothesis.

Unbiased Confidence Set

Since UMA sets rarely exist for two-sided tests, an alternative principle is **unbiasedness**.

Definition: A $(1-\alpha)$ confidence set $C(X)$ is **unbiased** if:

$$P_\theta\big(\theta' \in C(X)\big) \leq 1-\alpha \text{ for all } \theta \neq \theta'$$

- In other words, the interval doesn't "favor" wrong values over the true parameter.
- This parallels the definition of an **unbiased test**, where power under the alternative is always greater than size under the null.

3. Bayesian Optimality

Given a posterior distribution $\pi(\theta \mid x)$, the credible set $C(x)$ satisfies converge condition

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

Goal: Find $C(x)$ with the smallest size (length) among all sets with above probability.

Corollary: If $\pi(\theta \mid x)$ is unimodal, then the *shortest* credible interval for θ is:

$$C(x) = \{\theta\colon\ \pi(\theta \mid x) \geq k\},$$

where

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

This is called the Highest Posterior Density (HPD) region.

Example (Poisson HPD Region)

- Suppose $X_1, \dots, X_n \sim \mathrm{Poisson}(\lambda)$.
- With a conjugate **Gamma prior** $\pi(\lambda)$, the posterior is also Gamma.
- For prior $\mathrm{Gamma}(a, b)$, the posterior is:

$$\lambda \mid \sum x \sim \mathrm{Gamma}\left(a + \sum x,\ \frac{1}{n + 1/b}\right).$$

- The HPD credible region is:

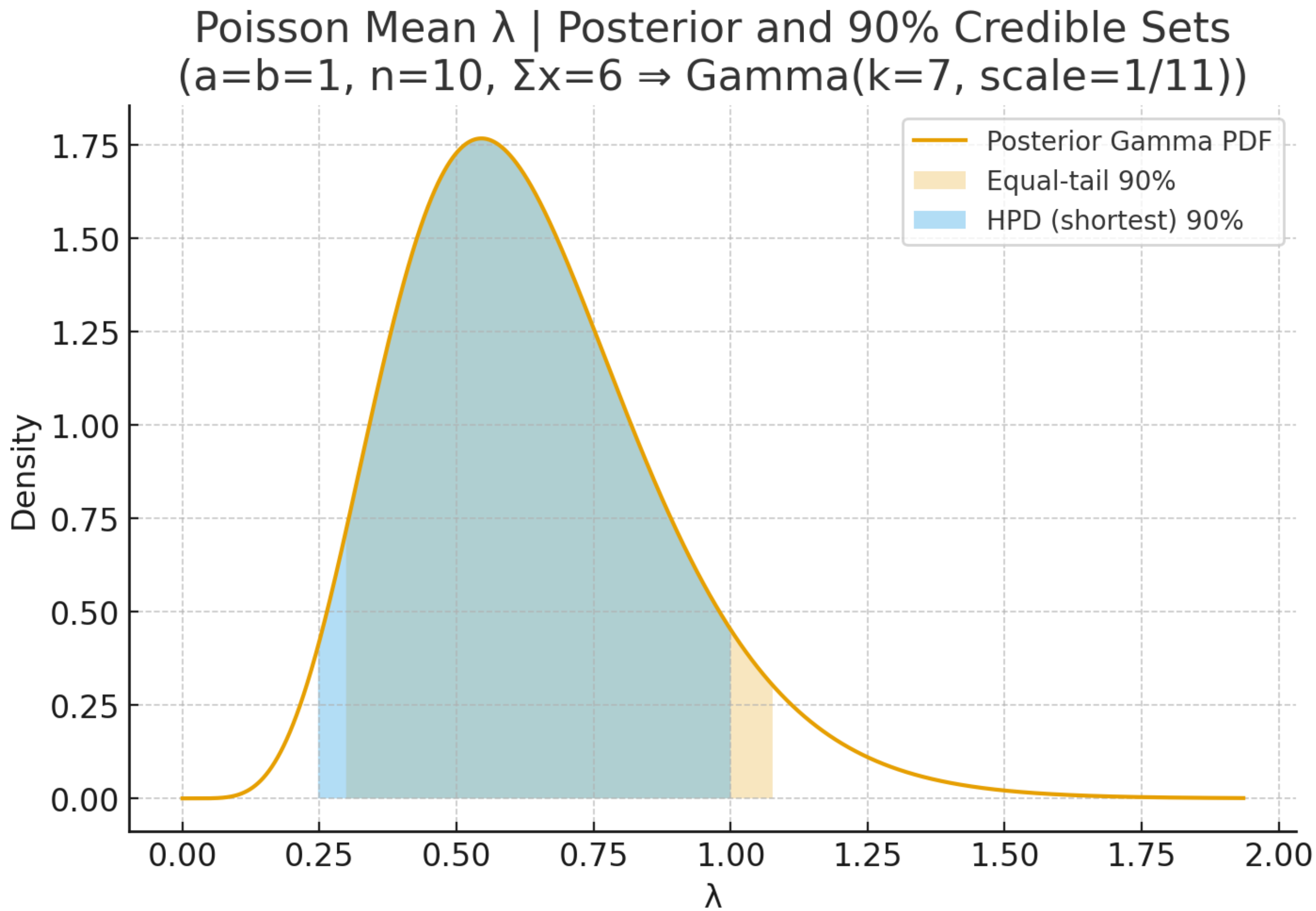
$$\{\lambda : \pi(\lambda \mid \sum x) \geq k\}, \quad \text{with} \int_{\{\lambda:\pi(\lambda| \sum x)\geq k\}} \pi(\lambda| \sum x)\, d\lambda = 1 - \alpha.$$

- In the specific case $a = b = 1, n = 10, \sum x = 6$:

  The 90% HPD credible set is approximately

$$[0.253,\ 1.005].$$

https://colab.research.google.com/drive/1rp2Xd3-YN7uUa8EBYS9gRQsDLGkYi-6f?usp=sharing



Interval	Lower	Upper	Length
Equal-tail 90%	0.299	1.077	0.778
HPD (shortest) 90%	0.247	1	0.753

4 Loss Function Optimality

Loss function optimality combines coverage and length into a single criterion.

Action space: "choosing a confidence set C."

Choose a rule $C(X)$ that minimizes expected loss.

Correctness: use indicator $I_C(\theta) = \begin{cases} 1, & \theta \in C, \\ 0, & \theta \notin C. \end{cases}$

One simple choice of Loss Function

$$L(\theta, C) := b \cdot Length(X) - I_C(\theta)$$

where $b > 0$ balances the trade-off:

Large b: prioritize shorter intervals.
Small b: prioritize coverage.

The **risk** is the expected loss under the sampling distribution:

$$R(\theta, C) = bE_\theta[Length(C(X))] - P_\theta(\theta \in C(X)).$$

So risk combines:

- Expected length (we want this small).
- Coverage probability (we want this large).

Example:

Suppose $X \sim N(\mu, \sigma^2)$,with σ^2known.

Define class of symmetric intervals:

$$C(X) = [X - c\sigma, X + c\sigma], c \geq 0.$$

Length: $\text{Len}(C) = 2c\sigma$.

Coverage $$P\big(\mu \in C(X)\big) = P\left(-c \leq \frac{X-\mu}{\sigma} \leq c\right) = 2\Phi(c) - 1$$

The risk is:

$$R(\mu, C) = b(2c\sigma) - (2\Phi(c) - 1).$$

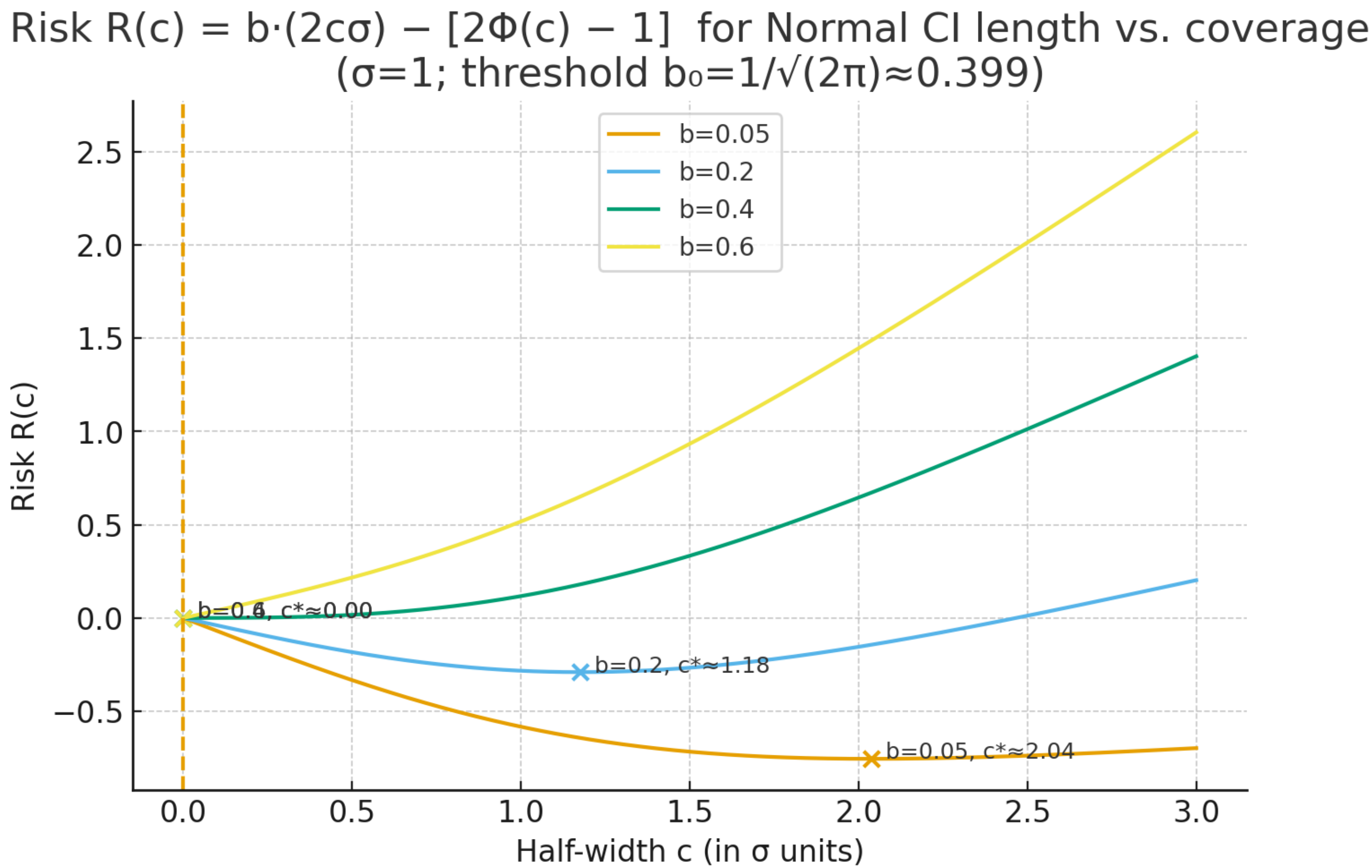
Minimizing Risk:

If $b\sigma > 1/\sqrt{2\pi}$,minimizing risk gives $c = 0$. Best estimator is **point estimator**.

If $b\sigma \leq 1/\sqrt{2\pi}$,minimizing risk gives

$$c = \sqrt{-2\ln\left(b\sigma\sqrt{2\pi}\right)}$$

If we express c as $z_{\alpha/2}$, the CI minimizes the risk corresponds to a standard confidence interval with confidence level $1 - \alpha$.



Each curve corresponds to a different b.
The markers show the minimizing c^* for that b.

References:

- **Book 1. [CB] Statistical Inference**, by Casella, George, Berger, Roger L, 2nd edition (Chapter 9.3)
- **Book 2. [W]: All of Statistics: Larry Wasserman**
- https://www.probabilitycourse.com/

Online books and courses:

- https://online.stat.psu.edu/stat415/
- https://stat110.hsites.harvard.edu/
- https://bookdown.org/egarpor/inference/