❖ **Interval Estimation 1**

-Finding Interval Estimators

Instructor: He Wang

Department of Mathematics

Northeastern University

Interval estimation

Sampling $X = \{X_1, \ldots, X_n\}$ from a population distribution with pdf $f(x|\theta)$

Knowledge of θ or $g(\theta)$ yields knowledge of the entire population.

Point Estimator $W(X_1, \ldots, X_n)$ estimates an unknown (fixed) parameter θ with a single number.

Definition: An interval estimator of a real parameter θ is a pair of functions $L(X)$ and $U(X)$ (with $L(X) \leq U(X)$), such that, with observed sample $(X = x)$

$$L(x) \leq \theta \leq U(x)$$

The random interval $[L(X), U(X)]$ is the estimator.

Example: Normal distribution.

Suppose random sample $X_i \sim Normal(\mu, \sigma^2)$ with known σ^2 for $i = 1, \ldots, n$

An interval estimator for μ is

$$\left[\bar{X} - k\frac{\sigma}{\sqrt{n}}, \bar{X} + k\frac{\sigma}{\sqrt{n}} \right]$$

for a chosen constant k. (For example, $k = z_{1-\alpha/2}$)

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

$$P\left(\mu \in \left[\bar{X} - k\tfrac{\sigma}{\sqrt{n}}, \ \bar{X} + k\tfrac{\sigma}{\sqrt{n}} \right] \right) = P\left(-k \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq k \right)$$
$$= P(-k \leq Z \leq k) \quad (Z \sim \mathcal{N}(0,1))$$
$$= 2\Phi(k) - 1.$$

Remark: $\bar{X}$ is the random variable. μ is an unknown constant!

Coverage probability

For an interval estimator $[L(X), U(X)][$ of a parameter θ, the **coverage probability at θ** is

$$P_\theta(\theta \in [L(X), U(X)]) = P_\theta(L(X) \le \theta \le U(X)).$$

The **confidence coefficient** (the guaranteed **coverage**) is the worst–case coverage over all θ :

$$\inf_\theta P_\theta(\theta \in [L(X), U(X)]).$$

We often say "**confidence interval**" for an interval estimator together with a target **coverage** (e.g., $1 - \alpha$)

Example: Normal distribution.

Suppose random sample $X_i \sim Normal(\mu, \sigma^2)$ with known σ^2 for $i = 1, \dots, n$

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

- Location-equivariant (good) intervals

Take fixed constants $c < d,$ an interval estimator for μ is

$$\left[\bar{X} + c\frac{\sigma}{\sqrt{n}}, \bar{X} + d\frac{\sigma}{\sqrt{n}} \right]$$

Coverage

$$P_\mu(\mu \in I_1(X)) = P_\mu\left(c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq d \right) = P(c \leq Z \leq d)$$
$$= \Phi(d) - \Phi(c).$$

This does not depend on μ, so the confidence coefficient equals $\Phi(d) - \Phi(c)$.

Special case (two-sided $1 - \alpha$ CI): choose $c = -z_{1-\alpha/2}$, $d = z_{1-\alpha/2}$, giving coverage 1−α and interval

$$\left[\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

- Multiplicative intervals (bad for a location family)

Consider $0 < a < 1 < b$ and

$$I_2(X) = [a\bar{X}, b\bar{X}].$$

Coverage at $\mu = 0$:

$$P_0\left(0 \in [a\bar{X}, b\bar{X}] \right) = P_0(a\bar{X} \leq 0 \leq b\bar{X}).$$

Since $a > 0$ and $b > 0$, the two inequalities force $\bar{X} \leq 0 \; and \; \bar{X} \geq 0$ simultaneously, i.e. $\bar{X} = 0$, an event of probability 0.

Methods for Building Interval Estimators

1. Inverting a test statistic
2. Pivotal quantities
3. Pivoting the CDF
4. Bayesian Intervals (Credible Intervals)

1. Inverting a test statistic

Example (two-sided normal, σ^2 known).

Suppose we have iid data $\mathcal{D} = \{x^{(1)}, \dots, x^{(n)}\}$ observed from normal distribution $N(\mu, \sigma^2)$ with known σ^2.

Test $H_0: \mu = \mu_0 \ v.s. \ H_1: \mu_1 \neq \mu_0$

At level α, the usual (UMP unbiased) test reject H_0 when

$$\frac{|\bar{X} - \mu_0|}{\sigma / \sqrt{n}} \geq z_{\alpha/2}$$

Equivalently, it **accepts** H_0 when

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Because this is a $size - \alpha$ test,

$$P_{\mu_0}(\text{accept } H_0) = 1 - \alpha.$$

But that probability statement holds for **every** μ_0. Therefore,

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Thus, **inverting** the acceptance region yields the familiar $(1 - \alpha)$ confidence interval

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Acceptance region vs confidence set

Acceptance region for $H_0: \mu = \mu_0$:

$$A(\mu_0) = \left\{ x \mid \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

Confidence set for μ given data x:

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

Theorem: (**Tests**↔**CI**)

- For each θ_0, let $A(\theta_0)$ be the **acceptance region** of a $level-\alpha$ test of $H_0: \theta = \theta_0$.

Define: $C(x) := \{\theta_0 \mid x \in A(\theta_0)\}$

Then, $C(X)$ is a $(1-\alpha)$ confidence set:

$$P_\theta\big(\theta \in C(X)\big) = P_\theta\big(x \in A(\theta)\big) \geq 1 - \alpha$$

- **Conversely**, given any $(1-\alpha)$ confidence set $C(X)$

Define $A(\theta_0) = \{x \mid \theta_0 \in C(x)\}$

Then $A(\theta_0)$ is the acceptance region of a level-α test of $H_0: \theta = \theta_0$.

Idea of proof. The first direction uses $P_\theta\big(x \in A(\theta)\big) \geq 1 - \alpha$. The second uses $P_\theta\big(\theta \notin C(X)\big) \leq \alpha$.

2. Pivotal quantities

A *pivot (pivotal quantity)* is a function $Q(\vec{X}, \theta)$ of the data $\vec{X}$ and parameter θ whose distribution does **not** depend on any unknown parameter θ.

That is, when $\vec{X} \sim F(\vec{x}|\theta)$, the distribution $Q(\vec{X}, \theta)$ is the same for every θ.

For any set $\mathcal{A}$ with $P_\theta\big(Q(\vec{X}, \theta) \in \mathcal{A}\big) = 1 - \alpha$, the set of parameter values

$$C(x) := \big\{\theta \mid Q(\vec{X}, \theta) \in \mathcal{A}\big\}$$

is a $(1 - \alpha)$ confidence set.

Pivotal quantities are at the heart of many classical confidence interval constructions.

Example (two-sided normal, σ^2 known).

Suppose we have iid data $\mathcal{D} = \{x^{(1)}, \ldots, x^{(n)}\}$ observed from normal distribution $N(\mu, \sigma^2)$ with known σ^2.

Sample Mean Estimator: $\bar{X}$

Pivotal quantity:

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

Use standard normal quantiles:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq z_{\alpha/2} \right) = 1 - \alpha.$$

Solve for μ

$$\mu \in \left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \ \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Example (Variance of Normal Distribution)

- Data: $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$.
- Estimator: Sample variance S^2.
- Pivotal quantity:

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

- Use chi-square quantiles:

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1 - \alpha.$$

- CI:

$$\sigma^2 \in \left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right].$$

Example: Proportion (Approximate Pivot)

- Data: $X \sim \text{Binomial}(n, p)$.
- Estimator: $\hat{p} = X/n$.
- Approximate pivot (CLT):

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

- This leads to the Wald confidence interval:

$$p \in \left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \; \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right].$$

3. Pivoting the CDF

Suppose we have a statistic $T(X)$ whose CDF depends on the parameter θ.

$$F_T(t; \theta) = P_\theta(T < t)$$

Find two functions $a(\theta), b(\theta)$ such that:

$$P\big(a(\theta) \leq T \leq b(\theta)\big) = 1 - \alpha$$

For observed data $T = t_0$, solve the inequalities for θ.

The solution yields the confidence interval for θ.

This method is useful when a pivotal quantity is not obvious, but the distribution (CDF) of an order statistic or sufficient statistic is known.

Example:

Suppose $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$, and we want CI for θ.

Statistic: $M = \max_i(X_i)$.

$$F_M(m; \theta) = P_\theta(M < m) = \left(\frac{m}{\theta}\right)^n \text{ for } 0 < m < \theta$$

Pick α split into tails:

$$P\left(M \geq \theta(1 - \alpha)^{1/n}\right) = \alpha$$

Solve θ, for observed $M = m$,

$$P\left(m \leq \theta \leq \frac{m}{(1 - \alpha)^{1/n}}\right) = \alpha$$

The confidence interval for θ *is* $\left[m, \frac{m}{(1 - \alpha)^{1/n}}\right]$

Example: Exponential Distribution

Suppose $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$, with mean $1/\lambda$

Statistic: $Y = \sum_i (X_i) \sim Gamma(n, \lambda)$.

Equivalently, $2\lambda Y \sim \chi^2_{2n}$

$$P\left(\chi^2_{2n,\alpha/2} \leq 2\lambda Y \leq \chi^2_{2n,1-\alpha/2}\right) = 1 - \alpha.$$

Solve λ,

$$\frac{\chi^2_{2n,\alpha/2}}{2Y} \leq \lambda \leq \frac{\chi^2_{2n,1-\alpha/2}}{2Y}$$

CI for mean lifetime $\theta = 1/\lambda$

$$\theta \in \left[\frac{2Y}{\chi^2_{2n,1-\alpha/2}}, \frac{2Y}{\chi^2_{2n,\alpha/2}}\right]$$

4. Bayesian Intervals (Credible Intervals)

Confidence intervals in the classical (frequentist) setting, we emphasize that the **interval is random**, while the parameter θ is **fixed**.

The randomness comes from the data, not from the parameter.

Given observed data, if we get 90% confidence interval for θ is [3,10], is is wrong in the frequentist sense to say:

"There is a 90% probability that θ lies in [3, 10]."

In Bayesian statistics, parameters are treated as **random variables** with a prior distribution θ and a posterior distribution given observed data $\theta|data$.

Thus, we can legitimately say:

"There is a 90% posterior probability that θ lies in [3, 10]."

General Procedure

Choose a prior distribution for θ, $\pi(\theta)$.

Suppose $\pi(\theta \mid \vec{x})$ is the posterior distribution of θ given data $\vec{X}$.

$$\pi(\theta \mid \vec{x}) = \frac{f(\vec{x}|\theta)\pi(\theta)}{\int f(\vec{x}|\theta')\pi(\theta')d\theta'}$$

Then for any set $A \subset \Theta$ the **credible probability** of A is:

$$P(\theta \in A|\vec{x}) = \int_A \pi(\theta|\vec{x})d\theta$$

For example, find **interval** [a,b] such that posterior probability mass inside equals $1 - \alpha$:

$$\int_a^b \pi(\theta|\vec{x})d\theta = 1 - \alpha$$

Example: Binomial Data with Beta Prior

Suppose $X \sim Binomial(n, p)$.

Prior: $p \sim Beta(\alpha, \beta)$.

Posterior:
$$p|X = x \sim Beta(\alpha + x, \beta + n - x)$$

A $100(1 - \alpha)\%$ credible interval for p is given by the $\alpha/2$ and $1 - \alpha/2$ quantiles of the Beta posterior distribution.
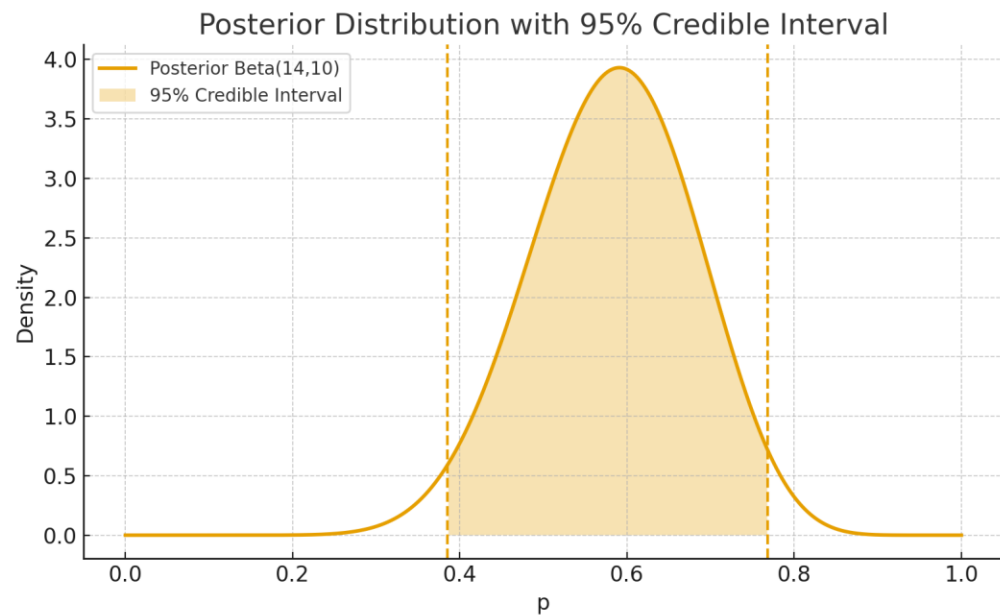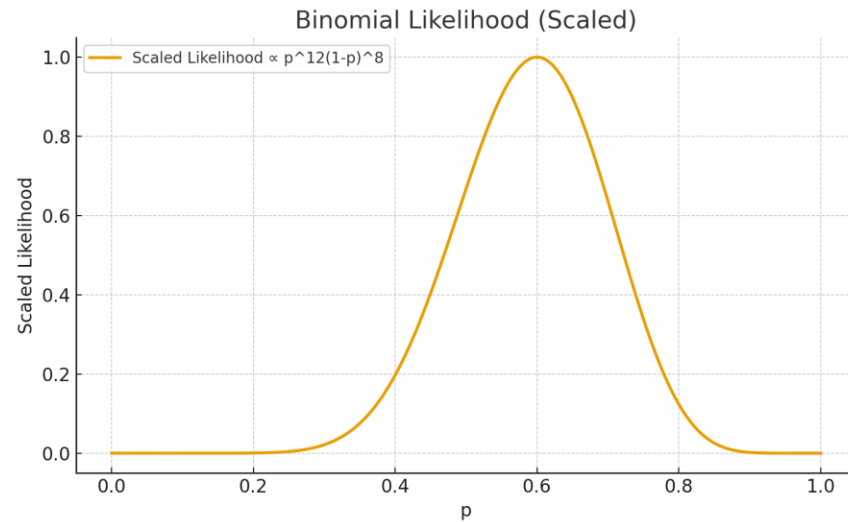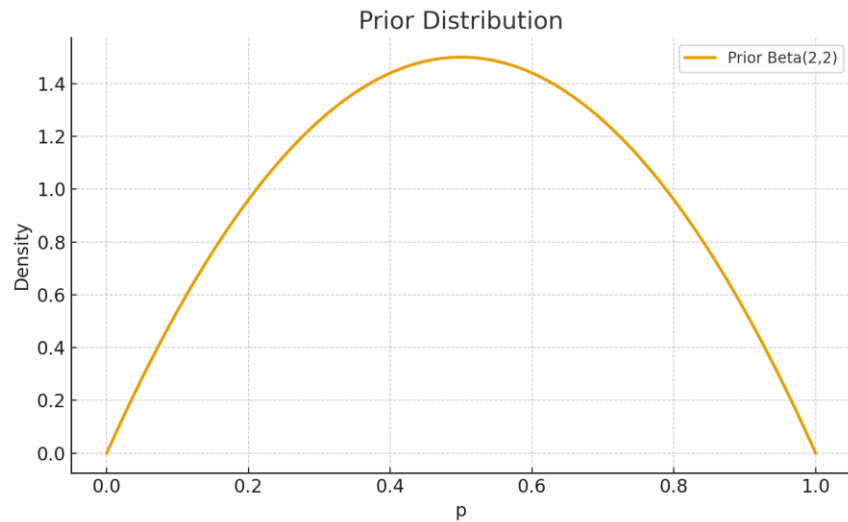
Suppose we conduct an experiment:
Number of trials: $n = 20$
Number of successes: $x = 12$
Prior: $p \sim Beta(2,2)$ (a weakly informative prior, centered at 0.5)

Posterior: $p \mid X = 12 \sim Beta(\alpha + x, \beta + n - x) = Beta(14,10)$

Prior Distribution

Density vs p

— Prior Beta(2,2)

Binomial Likelihood (Scaled)

Scaled Likelihood vs p

— Scaled Likelihood $\propto p^{12}(1-p)^8$

Posterior Distribution with 95% Credible Interval

Density vs p

— Posterior Beta(14,10)

95% Credible Interval

$p \in [0.385, 0.768]$

Normal Mean with Known Variance

Suppose we have iid data $\mathcal{D} = \{x^{(1)}, \ldots, x^{(n)}\}$ observed from normal distribution $N(\mu, \sigma^2)$ with known σ.

$$p(x^{(i)}|\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(x^{(i)} - \mu\right)^2\right)$$

The **prior** distribution of μ is $Normal(\theta, \tau^2)$

$$p(\mu) = \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{1}{2\tau^2}(\mu - \theta)^2\right)$$

The **posterior** of $\mu|\vec{x}$ is normal with mean and variance:

$$E(\mu|\vec{x}) = \frac{\tau^2\left(\sum_{i=1}^N x^{(i)}\right) + \sigma^2\theta}{n\tau^2 + \sigma^2} \qquad Var(\mu|\vec{x}) = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}$$

A $100(1 - \alpha)\%$ credible interval is the posterior mean $\pm z_{\alpha/2}$ times the posterior standard deviation.

Example: Poisson

Data: $\mathcal{D} = \{x^{(1)}, \ldots, x^{(n)}\} \sim Poisson(\lambda)$

Prior: $\lambda \sim Gamma(\alpha, \beta)$

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0.$$

Posterior: $\lambda \mid \vec{X} \sim Gamma(\alpha + \sum x^{(i)}, \beta + n)$.

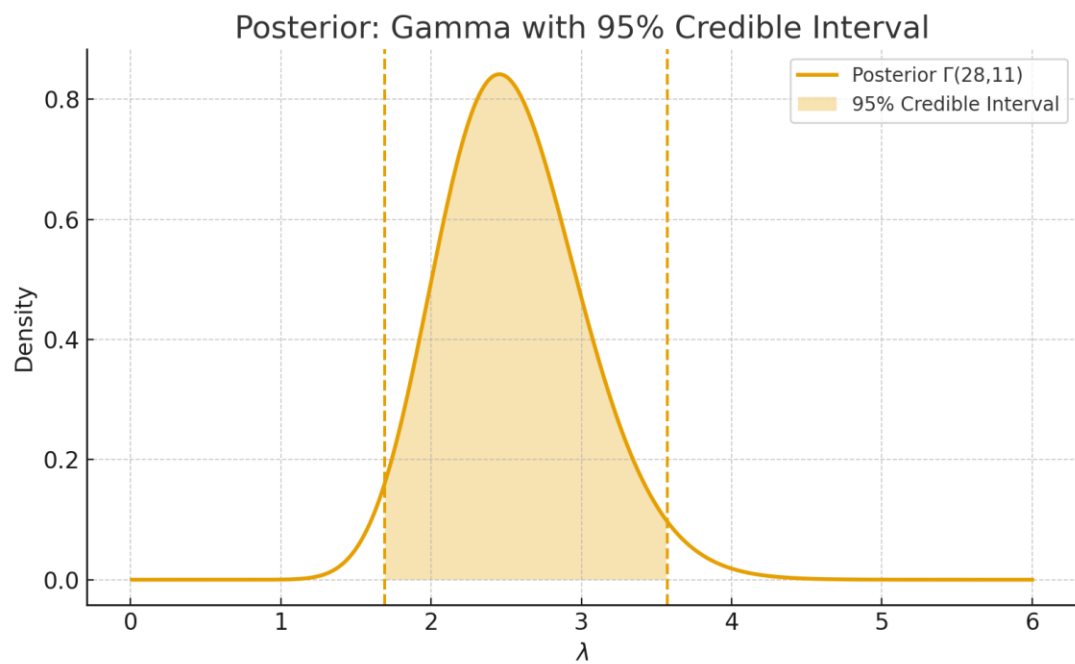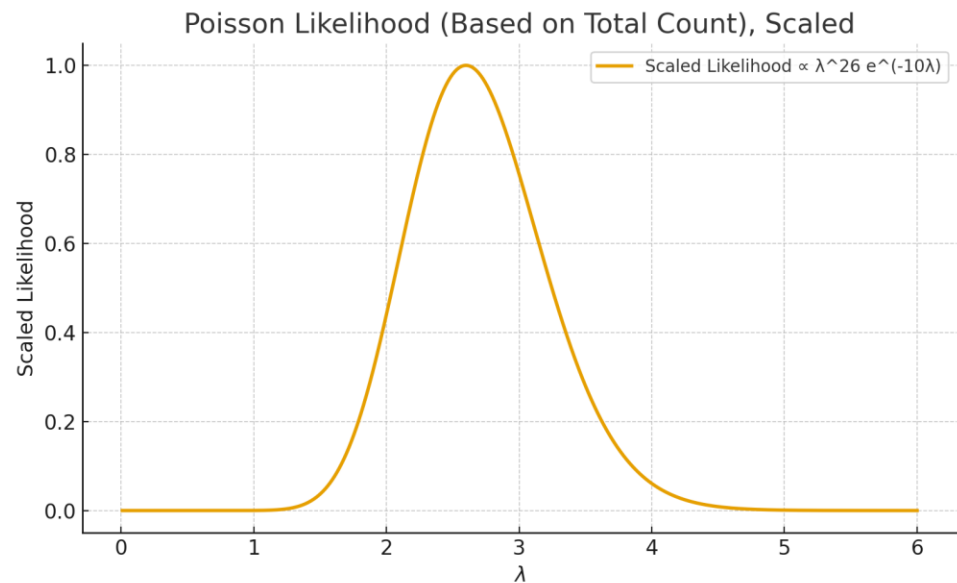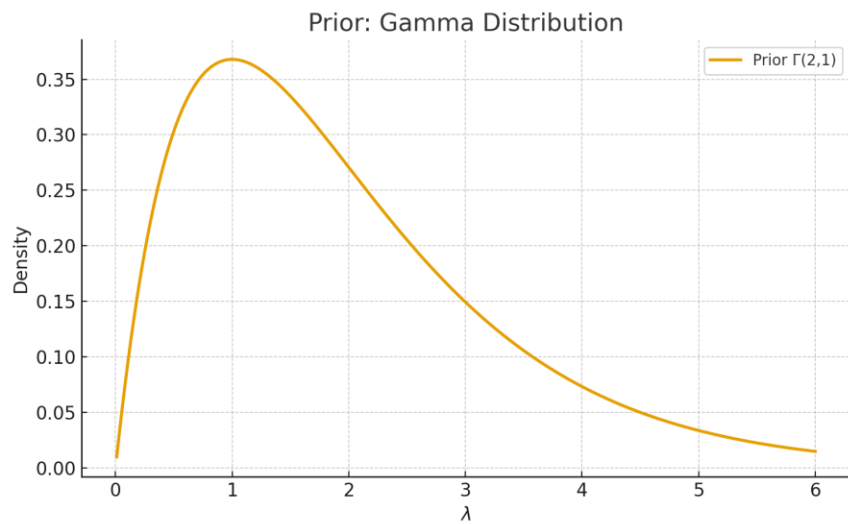Credible interval = quantiles of the Gamma posterior distribution.

For example:

n=10
Observed counts: $x_1, \ldots, x_{10}$ sum to $\sum x_i = 26$
Prior: $\lambda \sim Gamma(\alpha = 2, \beta = 1)$

Posterior: $\lambda \mid \vec{X} \sim Gamma(2 + 26, 1 + 10)$.

Prior: Gamma Distribution

Prior Γ(2,1)

Poisson Likelihood (Based on Total Count), Scaled

Scaled Likelihood ∝ λ^26 e^(-10λ)

Posterior: Gamma with 95% Credible Interval

Posterior Γ(28,11)

95% Credible Interval

Remarks

- Constructing and interpreting Bayesian credible intervals is generally more straightforward than frequentist confidence intervals.

- However, this simplicity comes at the cost of additional assumptions: one must specify a prior distribution.

- The Bayesian credible set depends on both the data and the chosen prior, unlike the frequentist confidence set, which depends only on the sampling distribution.

References:

- **Book 1. [CB] Statistical Inference**, by Casella, George, Berger, Roger L, 2nd edition (Chapter 8.3)

- **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/