**Exponential Distribution: Fisher Information**

To find the Fisher information I(θ) for a given distribution, we first need to calculate the log-likelihood function and its derivatives with respect to the parameter θ. The Fisher information is defined as the negative expectation of the second derivative of the log-likelihood function with respect to θ, or the expectation of the square of the first derivative.

Given X ~ exp(θ), with the probability density function f(x, θ) = (1/θ) \* exp(-x/θ), where θ > 0 and x > 0, let's find the log-likelihood function L(θ).

The log-likelihood function is given by: L(θ) = log(f(x, θ)) = log((1/θ) \* exp(-x/θ)) = -log(θ) - x/θ

Now, let's find the first and second derivatives of L(θ) with respect to θ:

First derivative: dL(θ)/dθ = -1/θ + x/θ^2

Second derivative: d²L(θ)/dθ² = 1/θ² - 2x/θ³

Now, we need to find the Fisher information I(θ), which is given by:

I(θ) = E[-d²L(θ)/dθ²] = -E[1/θ² - 2x/θ³]

Since X ~ exp(θ), we know that E[X] = θ. Therefore, we can substitute E[X] with θ:

I(θ) = -(-1/θ² + 2θ/θ³) = 1/θ²

For the observed information i\_n(θ), we first need to find the average of the second derivative of the log-likelihood function over the sample. Let's assume we have a sample of n observations: x\_1, x\_2, ..., x\_n. The log-likelihood function for the entire sample is:

L\_n(θ) = Σ(-log(θ) - x\_i/θ) for i = 1 to n

Taking the second derivative with respect to θ:

d²L\_n(θ)/dθ² = Σ(1/θ² - 2x\_i/θ³) for i = 1 to n

Now, the observed information i\_n(θ) is the negative of the average of the second derivative:

i\_n(θ) = -1/n \* d²L\_n(θ)/dθ² = -1/n \* Σ(1/θ² - 2x\_i/θ³) for i = 1 to n

Discuss and give math functions for Cramer - Rao Lower bound

The Cramér-Rao Lower Bound (CRLB) is a fundamental result in statistical estimation theory that provides a lower bound on the variance of any unbiased estimator for a given parameter. In other words, it gives us a benchmark for the best possible precision that an unbiased estimator can achieve.

The Cramér-Rao Lower Bound is derived from the Fisher information and is given by the following inequality:

Var(θ̂) ≥ 1 / I(θ)

where Var(θ̂) represents the variance of the unbiased estimator θ̂ of the parameter θ, and I(θ) is the Fisher information for the parameter θ.

The CRLB states that no unbiased estimator can have a variance lower than the reciprocal of the Fisher information. This implies that the more information available about the parameter, the lower the variance of the best possible unbiased estimator.

To derive the CRLB, we first consider any unbiased estimator θ̂ for the parameter θ, which satisfies:

E[θ̂] = θ

Now, let's consider the score function, which is the first derivative of the log-likelihood function with respect to the parameter θ:

U(θ) = dL(θ)/dθ

By definition, the Fisher information is the expected value of the square of the score function:

I(θ) = E[U(θ)^2]

Now, consider the difference between the score function and its expected value:

V(θ) = U(θ) - E[U(θ)]

The key observation is that V(θ) and θ̂ are uncorrelated, meaning E[V(θ) \* θ̂] = 0. This can be shown using the properties of the log-likelihood function and the unbiasedness of the estimator.

Using this result, we can express the variance of the estimator as:

Var(θ̂) = E[(θ̂ - θ)^2] = E[(θ̂ - E[θ̂] + V(θ))^2] = E[V(θ)^2] + 2E[V(θ) \* θ̂] + E[θ̂^2]

Since E[V(θ) \* θ̂] = 0, we have:

Var(θ̂) = E[V(θ)^2] + E[θ̂^2] - θ^2

Now, we use the Cauchy-Schwarz inequality, which states that for any two random variables X and Y:

(E[X \* Y])^2 ≤ E[X^2] \* E[Y^2]

Applying the Cauchy-Schwarz inequality to V(θ) and θ̂, we get:

E[U(θ) \* θ̂]^2 ≤ E[U(θ)^2] \* E[θ̂^2]

However, E[U(θ) \* θ̂] = I(θ) \* (E[θ̂^2] - θ^2), so:

I(θ) \* (E[θ̂^2] - θ^2) ≤ E[U(θ)^2] \* E[θ̂^2]

Dividing both sides by E[θ̂^2], we obtain:

I(θ) \* (1 - θ^2 / E[θ̂^2]) ≤ E[U(θ)^2]

Since E[U(θ)^2] = I(θ), we have:

I(θ) \* (1 - θ^2 / E[θ̂^2]) ≤ I(θ)

Rearranging the terms and using the

continue

Rearranging the terms and using the fact that I(θ) > 0, we get:

1 - θ^2 / E[θ̂^2] ≤ 1

This implies:

θ^2 / E[θ̂^2] ≥ 0

Hence, we have:

E[θ̂^2] - θ^2 ≥ 0

Now, recall that Var(θ̂) = E[θ̂^2] - θ^2. Therefore, we can write:

Var(θ̂) ≥ 0

Using the previous result that E[U(θ)^2] = I(θ), we can express the Cramér-Rao Lower Bound as:

Var(θ̂) ≥ 1 / I(θ)

This inequality states that the variance of any unbiased estimator θ̂ of the parameter θ is greater than or equal to the reciprocal of the Fisher information for that parameter. In other words, the CRLB gives a lower limit on the precision that an unbiased estimator can achieve, and it shows that the more information we have about the parameter (as measured by the Fisher information), the smaller the variance of the best possible unbiased estimator.

**Efficiency of S^2**

The efficiency of an estimator is a measure of its performance compared to the Cramér-Rao Lower Bound (CRLB), which is the best possible variance an unbiased estimator can achieve. An estimator is said to be efficient if it reaches the CRLB. Let's examine the efficiency of the sample variance (S²) as an estimator of the population variance (σ²) for a normally distributed sample.

Given a sample X₁, X₂, ..., Xₙ from a normal distribution with mean μ and variance σ², the sample variance S² is defined as:

S² = Σ((Xᵢ - X̄)²) / (n - 1)

where X̄ is the sample mean:

X̄ = Σ(Xᵢ) / n

Now, let's find the efficiency of S² as an estimator for σ². The first step is to find the Fisher information I(σ²) for the normal distribution.

For a normal distribution, the probability density function is given by:

f(x, μ, σ²) = (1 / √(2πσ²)) \* exp(-(x - μ)² / (2σ²))

The Fisher information I(σ²) for the variance σ² can be computed using the log-likelihood function and its second derivative with respect to σ². However, for the normal distribution, it's easier to work with the Fisher information for the precision τ, which is the reciprocal of the variance (τ = 1 / σ²). The Fisher information for τ is given by:

I(τ) = n / 2τ²

Since τ = 1 / σ², the Fisher information for σ² can be derived as:

I(σ²) = n / (2σ⁴)

Now we can find the CRLB for an unbiased estimator of σ²:

CRLB(σ²) = 1 / I(σ²) = 2σ⁴ / n

The next step is to calculate the variance of the sample variance S². It can be shown that for a normal distribution:

Var(S²) = (2σ⁴) / (n - 1)

Now, let's compare the variance of S² to the CRLB to determine the efficiency:

Efficiency = CRLB(σ²) / Var(S²) = (2σ⁴ / n) / (2σ⁴ / (n - 1))

Efficiency = (n - 1) / n

As n becomes large, the efficiency approaches 1, which means that the sample variance S² becomes an efficient estimator of the population variance σ² in large samples. However, for small sample sizes, the efficiency is less than 1, indicating that S² is not an efficient estimator of σ² in those cases.

**Asymptotic normality of MLEs**

Asymptotic normality is an important property of maximum likelihood estimators (MLEs) in the context of large sample theory. As the sample size (n) grows to infinity, the maximum likelihood estimator converges to the true parameter value, and its sampling distribution becomes approximately normal. This property is useful for understanding the behavior of MLEs in large samples and for constructing confidence intervals and hypothesis tests.

Let θ be the true parameter value, and let θ̂ be the maximum likelihood estimator of θ based on a sample of size n. The asymptotic normality of MLEs states that, under certain regularity conditions:

√n (θ̂ - θ) → N(0, I(θ)^(-1))

where "→" denotes convergence in distribution, N(0, I(θ)^(-1)) is a normal distribution with mean 0 and variance equal to the inverse of the Fisher information I(θ) evaluated at the true parameter value θ, and √n is the square root of the sample size.

The asymptotic normality of MLEs relies on the following regularity conditions:

1. The parameter space is a compact subset of the real line, or the likelihood function is bounded.
2. The true parameter value θ lies in the interior of the parameter space.
3. The likelihood function is twice continuously differentiable with respect to θ.
4. The expected value of the score function (the first derivative of the log-likelihood function) is 0: E[U(θ)] = 0, where U(θ) = dL(θ)/dθ.
5. The Fisher information I(θ) is positive and finite for all θ in the parameter space.
6. The log-likelihood function satisfies some regularity conditions, such as the dominated convergence theorem.

When these conditions are satisfied, the maximum likelihood estimator θ̂ is asymptotically normal, meaning that its distribution approaches a normal distribution with mean θ and variance I(θ)^(-1)/n as the sample size n goes to infinity.

The asymptotic normality of MLEs has important implications for statistical inference:

1. Confidence intervals: Asymptotic normality allows us to construct approximate confidence intervals for the true parameter value θ using the MLE θ̂ and the estimated standard error based on the Fisher information.
2. Hypothesis testing: Asymptotic normality enables us to conduct hypothesis tests for the true parameter value θ by comparing the MLE θ̂ to a normal distribution with mean θ under the null hypothesis and the estimated standard error.
3. Asymptotic efficiency: Asymptotic normality combined with the Cramér-Rao lower bound implies that MLEs are asymptotically efficient, meaning they achieve the lowest possible variance among unbiased estimators in large samples.

**Derive asymptotic normality**

Asymptotic normality of maximum likelihood estimators (MLEs) is an essential property in large sample theory. It states that the distribution of the MLE approaches a normal distribution with mean equal to the true parameter value and variance equal to the inverse of the Fisher information as the sample size tends to infinity. In this derivation, we will focus on a one-dimensional parameter θ, although the concept can be extended to multidimensional parameters.

Let's derive the asymptotic normality of MLEs under certain regularity conditions:

1. The parameter space is a compact subset of the real line, or the likelihood function is bounded.
2. The true parameter value θ lies in the interior of the parameter space.
3. The likelihood function is twice continuously differentiable with respect to θ.
4. The expected value of the score function (the first derivative of the log-likelihood function) is 0: E[U(θ)] = 0, where U(θ) = dL(θ)/dθ.
5. The Fisher information I(θ) is positive and finite for all θ in the parameter space.
6. The log-likelihood function satisfies some regularity conditions, such as the dominated convergence theorem.

Given a random sample X₁, X₂, ..., Xₙ, let L(θ) denote the log-likelihood function, and U(θ) denote the score function (first derivative of L(θ) with respect to θ). The Fisher information I(θ) is the expected value of the square of the score function:

I(θ) = E[U(θ)²]

According to the Taylor series expansion, we can write the score function around the true parameter value θ as:

U(θ̂) = U(θ) + (θ̂ - θ) \* U'(θ) + O((θ̂ - θ)²)

where θ̂ is the MLE, U'(θ) is the second derivative of L(θ) with respect to θ, and O((θ̂ - θ)²) denotes the higher-order terms.

Since the MLE θ̂ maximizes the log-likelihood function, the score function evaluated at θ̂ is equal to 0:

U(θ̂) = 0

Therefore, we have:

(θ̂ - θ) \* U'(θ) = -U(θ) + O((θ̂ - θ)²)

Now, let's multiply both sides by the square root of the sample size (√n) and rearrange the terms:

√n (θ̂ - θ) = -√n \* U(θ) / U'(θ) + √n \* O((θ̂ - θ)²)

By the Central Limit Theorem (CLT), √n \* U(θ) converges in distribution to a normal distribution with mean 0 and variance I(θ). Furthermore, as n goes to infinity, the term √n \* O((θ̂ - θ)²) converges to 0. Thus, we have:

√n (θ̂ - θ) → N(0, I(θ)^(-1))

This is the asymptotic normality result, which states that the distribution of the MLE approaches a normal distribution with mean equal to the true parameter value θ and variance equal to the inverse of the Fisher information I(θ) as the sample size tends to infinity.

**How is chebyshev inequality involved**

Chebyshev's inequality is a fundamental result in probability theory that provides a bound on the probability that a random variable deviates from its mean by a certain amount. Chebyshev's inequality states that for any random variable X with finite mean (μ) and variance (σ²):

P(|X - μ| ≥ kσ) ≤ 1/k²

where k > 0 is a constant.

While Chebyshev's inequality is not directly involved in the derivation of the asymptotic normality of MLEs, it can be used to illustrate the convergence of the MLE to the true parameter value. Specifically, it can help to show that the MLE converges to the true parameter value in probability as the sample size increases, which is known as consistency.

In the context of MLEs, let θ̂ be the MLE of the parameter θ based on a sample of size n, and let μ = θ be the true parameter value. Also, let σ² denote the variance of the MLE, which is typically equal to the inverse of the Fisher information I(θ) divided by the sample size n for large samples:

σ² ≈ I(θ)^(-1)/n

Now, using Chebyshev's inequality, we can bound the probability that the MLE deviates from the true parameter value by more than k times its standard deviation:

P(|θ̂ - θ| ≥ kσ) ≤ 1/k²

As the sample size n increases, the variance σ² decreases, which implies that the probability of the MLE deviating from the true parameter value by more than k times its standard deviation becomes smaller. In other words, as n goes to infinity, the MLE converges to the true parameter value in probability:

θ̂ → θ in probability as n → ∞

In summary, while Chebyshev's inequality is not directly involved in the derivation of the asymptotic normality of MLEs, it can be used to illustrate the consistency of the MLE by showing that the probability of the MLE deviating from the true parameter value by more than a certain amount decreases as the sample size increases.

**In a table, show the biased and unbiased estimators**

In this table, I will list some common biased and unbiased estimators for population parameters, specifically the population mean (μ), variance (σ²), and standard deviation (σ). Note that these estimators are based on a random sample from the population.

| **Population Parameter** | **Biased Estimator** | **Unbiased Estimator** |
| --- | --- | --- |
| Mean (μ) | X̄ = Σ(Xᵢ) / n | X̄ = Σ(Xᵢ) / n |
| Variance (σ²) | S²\_b = Σ((Xᵢ - X̄)²) / n | S²\_u = Σ((Xᵢ - X̄)²) / (n-1) |
| Standard Deviation (σ) | Not Applicable | Not Applicable |

1. Population Mean (μ):
   * Biased Estimator: The sample mean (X̄) is an unbiased estimator of the population mean (μ) because its expected value is equal to the population mean.
   * Unbiased Estimator: The sample mean (X̄) is also the unbiased estimator for the population mean (μ).
2. Population Variance (σ²):
   * Biased Estimator: The biased estimator for the population variance (S²\_b) is calculated by dividing the sum of squared differences between the sample values and the sample mean (X̄) by the sample size (n). This estimator is biased because its expected value is smaller than the population variance.
   * Unbiased Estimator: The unbiased estimator for the population variance (S²\_u) is calculated by dividing the sum of squared differences between the sample values and the sample mean (X̄) by (n - 1), where n is the sample size. This estimator is unbiased because its expected value is equal to the population variance.
3. Population Standard Deviation (σ):
   * Biased Estimator: There is no straightforward biased estimator for the population standard deviation. In practice, the square root of the biased variance estimator is sometimes used, but it does not have the same bias properties as the biased variance estimator.
   * Unbiased Estimator: There is no unbiased estimator for the population standard deviation that can be expressed in a simple closed form. In practice, the square root of the unbiased variance estimator (S = √(S²\_u)) is used as an approximation, but it is not strictly unbiased.

The above table provides examples of biased and unbiased estimators for some common population parameters. Note that the choice of an estimator depends on the specific problem and the properties desired for the estimator (e.g., bias, efficiency, robustness).