Point estimators

Probability theory. Statistics

Anton Afanasev

Higher School of Economics

DSBA 221

November 25, 2023

Seminar Overview

1. Quiz
2. Point estimators. Specific cases
   - Sample variance
   - Sample proportion
3. Point estimators. General case
   - Properties
   - Mean squared error
   - Consistency
4. Practice

Quiz

Let $Z_1, \dots, Z_7$ be a random sample from the standard normal distribution.
Let $W = Z_1^2 + \dots + Z_7^2$.

1. Use CLT to estimate $\mathsf{P}(1.69 < W < 14.07)$.
2. Find exact value of that probability.

Sample variance

- Let sample $X_1, \ldots, X_n$ from the same population be i.i.d. random variables with $\mathsf{E}(X_i) = \mu$ and $\mathsf{V}(X_i) = \sigma^2$. Mean μ is unknown.
- As in the example with sample mean, it would seem that the natural estimator for variance is:

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

- But it turns out that:

$$\mathsf{E}\left(\widehat{\sigma^2}\right) = \frac{n-1}{n}\sigma^2.$$

- That's why unbiased sample variance S^2 is used with $\mathsf{E}\left(S^2\right) = \sigma^2$:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

Degrees of freedom

Definition

Degrees of freedom in statistics – number of independent values, which can be varied without breaking any constraints, while estimating a parameter.

Example

- Let's consider a sample with size 10 and already estimated mean value:

$$2 \quad -1 \quad 4 \quad 9 \quad -2 \quad -2 \quad 3 \quad -2 \quad -6 \quad x$$
$$\text{Sum} = 10, \qquad \text{Mean} = 1.$$

- The last value has no freedom to vary, since it is tightly connected to known mean value. The only possible outcome is $x = 5$.
- Mean imposes a constraint on a freedom to vary.
- Thus, in this example a number of degrees of freedom is 9.

Degrees of freedom

- Sample variance has $n - 1$ degrees of freedom.
- Constraint:
$$\sum_{i=1}^{n} (X_i - \overline{X}) = 0.$$
- Sample variance is calculated from the vector of residuals $X_i - \overline{X}$:
$$\begin{pmatrix} X_1 - \overline{X} & X_2 - \overline{X} & \dots & X_n - \overline{X} \end{pmatrix}.$$
- While there are n independent observations in the sample, there are only $n - 1$ independent residuals, as they sum to 0.
- Overall, number of degrees of freedom is calculated as:
$$\text{DF} = \text{Sample size} - \text{Number of constraints}.$$

What if μ is known?

- If μ is known, we can use it to estimate σ:

$$\varsigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 .$$

- Such estimator has n degrees of freedom, since residuals have no constraints, and subsequently:

$$\mathsf{E}\left(\varsigma^2\right) = \sigma^2.$$

- We can easily derive a distribution for $\widehat{\sigma^2}$:

$$\frac{\varsigma^2}{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{n} \sum_{i=1}^{n} Z_i^2,$$

$$\frac{n\varsigma^2}{\sigma^2} \sim \chi^2_n.$$

Fisher's lemma

- Let $X_1, \ldots, X_n$ be i.i.d. with $X_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$.
- Distribution of $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ is given by:

Lemma (Fisher)

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

Problem 1

Suppose $X_1, X_2, \ldots, X_{10}$ is a random sample taken from a $\mathcal{N}(12, 25)$-distributed population.

1. Find the probability that the sample variance S^2 is between 20 and 30.
2. Find the range for the middle 90% of the distribution of the sample variance.

Problem 1

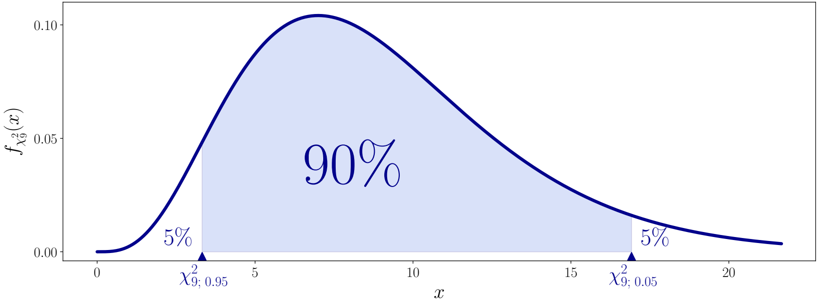

Figure: Middle 90% of χ^2_9-distribution.

Sample proportion

- Let X be a number of successes in n trials with probability of success p:
$$X \sim \mathrm{Bin}(n, p).$$
- Probability p can be estimated via sample proportion $\widehat{P}$:
$$\widehat{P} = \frac{X}{n}.$$
- Expected value and variance of $\widehat{P}$:
$$\mathsf{E}\left(\widehat{P}\right) = \mathsf{E}\left(\frac{X}{n}\right) = \frac{\mathsf{E}\left(X\right)}{n} = \frac{np}{n} = p,$$
$$\mathsf{V}\left(\widehat{P}\right) = \mathsf{V}\left(\frac{X}{n}\right) = \frac{\mathsf{V}\left(X\right)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$
- $\widehat{P}$ is NOT binomial, since it can take non-integer values.

Sample proportion

- Since explicit form of distribution for $\widehat{P}$ is unknown, let's use CLT (or an extension of De Moivre–Laplace theorem, scaled by n):
$$\frac{\widehat{P} - \mathsf{E}\left(\widehat{P}\right)}{\sigma\left(\widehat{P}\right)} = \frac{\widehat{P} - p}{\sqrt{p(1-p)/n}} \xrightarrow[n\to\infty]{d} Z \sim \mathcal{N}\left(0, 1\right).$$
- Thus:
$$\widehat{P} \overset{\text{CLT}}{\sim} \mathcal{N}\left(p, \frac{p(1-p)}{n}\right).$$
- Normal approximation of $\widehat{P}$ is limited by it's physical meaning – proportion can not exceed bounds $[0, 1]$. Since $\mathsf{E}\left(\widehat{P}\right) \pm 3 \cdot \sigma\left(\widehat{P}\right)$ contains 99.7% of all possible values of $\widehat{P}$:
$$\left[p - 3\sqrt{\frac{p(1-p)}{n}},\ p + 3\sqrt{\frac{p(1-p)}{n}}\right] \subset [0, 1],$$

Problem 2

An ordinary die is “fair” or “balanced” if each face has an equal chance of landing on top when the die is rolled. Thus the proportion of times a three is observed in a large number of tosses is expected to be close to 1/6. Suppose a die is rolled 240 times and shows three on top 36 times.

1. Find the probability that a fair die would produce a proportion of 0.15 or less.
2. Give an interpretation of the result in part 1. How strong is the evidence that the die is not fair?
3. Suppose the sample proportion 0.15 came from rolling the die 2,400 times instead of only 240 times. Rework part 1 under these circumstances.
4. Give an interpretation of the result in part 3. How strong is the evidence that the die is not fair?

Point estimators

- $X_1, \ldots, X_n$ – sample from population with size n, assumed to be i.i.d. random variables.
- X_i have common c.d.f. $F_{X_i}(x;\theta)$, where θ is a parameter of distribution.

Definition

Point estimator of parameter θ is an arbitrary function of a sample:

$$\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n).$$

- Point estimators have bias: $\operatorname{Bias}\left(\widehat{\theta}_n\right) = \mathsf{E}\left(\widehat{\theta}_n\right) - \theta$.

Definition

Point estimator is called unbiased if: $\mathsf{E}\left(\widehat{\theta}_n\right) = \theta.$

Risk functions

- Loss function $u\left(\widehat{\theta}_n - \theta\right)$ – penalizes the choice of point estimator, based on a "distance" between true parameter and its estimate.
- Properties of loss function:
  1. $u(0) = 0$,
  2. $u(-x) = x$,
  3. $u(x)$ is monotonous.
- Expected loss is characterized by risk function $R_{\widehat{\theta}_n}(\theta)$:

$$R_{\widehat{\theta}_n}(\theta) = \mathsf{E}\left(u\left(\widehat{\theta}_n - \theta\right)\right).$$

Example

- $\mathsf{E}\left|\widehat{\theta}_n - \theta\right|$ – mean absolute error (MAE),
- $\mathsf{E}\left(\widehat{\theta}_n - \theta\right)^2$ – mean squared error (MSE),

Mean squared error

- MSE of estimator θ can be represented as

$$\mathrm{MSE}\left(\widehat{\theta}_n\right) = \mathsf{E}\left(\widehat{\theta}_n - \theta\right)^2 = \mathsf{V}\left(\widehat{\theta}_n - \theta\right) + \left(\mathsf{E}\left(\widehat{\theta}_n - \theta\right)\right)^2 = \\ = \mathsf{V}\left(\widehat{\theta}_n\right) + \mathrm{Bias}^2\left(\widehat{\theta}_n\right).$$

- Minimizing MSE is a key criterion in selecting estimators.

Problem 3

Random variable X assumes values 0 and 1, each with probability $1/2$.

1. Find population mean μ and variance σ^2.
2. You have 9 independent observations of X: $\{X_1, \ldots, X_9\}$. Consider the following estimators of the population mean μ:

   1. $\widehat{\mu}_1 = 0.45$;
   2. $\widehat{\mu}_2 = X_1$;
   3. $\widehat{\mu}_3 = \overline{X}$;
   4. $\widehat{\mu}_4 = X_1 + \frac{1}{3}X_2$;
   5. $\widehat{\mu}_5 = \frac{2}{3}X_1 + \frac{2}{3}X_2 - \frac{1}{3}X_3$.

   Which of these estimators are unbiased? Calculate bias for each estimator. Which estimator is the most efficient?
3. Which estimators from part 2 are consistent?

Consistent estimators

Definition

Estimator $\widehat{\theta}_n$ is called consistent if:

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{\mathsf{P}} \theta.$$

or

$$\forall \varepsilon > 0 : \qquad \mathsf{P}\left(\left|\widehat{\theta}_n - \theta\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

Markov's inequality

If X is a non-negative random variable, then $\forall \varepsilon > 0$

$$\mathsf{P}(X \geq \varepsilon) \leq \frac{\mathsf{E}(X)}{\varepsilon}.$$

Consistent estimators

- Let's use $\left|\widehat{\theta}_n - \theta\right|$ in Markov's inequality:

$$\mathsf{P}\left(\left|\widehat{\theta}_n - \theta\right| \geq \varepsilon\right) = \mathsf{P}\left(\left(\widehat{\theta}_n - \theta\right)^2 \geq \varepsilon^2\right) \leq$$
$$\leq \frac{\mathsf{E}\left(\widehat{\theta}_n - \theta\right)^2}{\varepsilon^2} = \frac{\text{MSE}\left(\widehat{\theta}_n\right)}{\varepsilon^2}.$$

- Sufficient condition of consistency:

$$\text{MSE}\left(\widehat{\theta}_n\right) \xrightarrow[n\to\infty]{} 0.$$

Chebyshev's inequality

If X is a variable with $\mathsf{E}|X| < \infty$ and $0 < \mathsf{V}(X) < \infty$, then $\forall \varepsilon > 0$

$$\mathsf{P}\left(|X - \mathsf{E}(X)| > \varepsilon\right) \leq \frac{\mathsf{V}\left(X\right)}{\varepsilon^2}.$$

Problem 4

Let X_1, X_2, X_3 be a random sample from a population with mean μ and variance σ^2. Consider the following two estimators of variance σ^2:

1. $\widehat{\sigma_1^2} = c_1 \left(X_1 - X_2\right)^2$;
2. $\widehat{\sigma_2^2} = c_2 \left(X_1 - X_2\right)^2 + c_2 \left(X_1 - X_3\right)^2 + c_2 \left(X_2 - X_3\right)^2$.

Find constants c_1, c_2, such that $\widehat{\sigma_1^2}$ and $\widehat{\sigma_2^2}$ are unbiased estimators of σ^2.

Problem 5

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(2\mu, 2\sigma^2)$. You have samples of size n and m from the two distributions: $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$. Consider the estimator $\widehat{\mu} = c_1\overline{X} + c_2\overline{Y}$.

1. For which c_1, c_2 the estimator is unbiased?
2. For which c_1, c_2 the estimator is unbiased and most efficient?

Problem 6

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a $\mathcal{U}(0, \theta)$ distribution, where θ is unknown. Define the estimator

$$\widehat{\Theta}_n = \max\{X_1, X_2, X_3, \ldots, X_n\}.$$

1. Find the bias of $\widehat{\Theta}_n$.
2. Find the MSE of $\widehat{\Theta}_n$.
3. Is $\widehat{\Theta}_n$ a consistent estimator of θ?

Problem 7

When R successes occur in n trials, the sample proportion $\widehat{p} = R/n$ customarily is used as an estimator of the probability of success p. However, there are sometimes good reasons to use the estimator $p^* \equiv \dfrac{R+1}{n+2}$. Alternatively, p^* can be written as a linear combination of the familiar estimator $\widehat{p}$:

$$p^* = \frac{n\widehat{p}+1}{n+2} = \frac{n}{n+2} \cdot \widehat{p} + \frac{1}{n+2}.$$

1. What is the MSE of $\widehat{p}$? Is it consistent?
2. What is the MSE of p^* ? Is it consistent?
3. To decide which estimator is better, $\widehat{p}$ or p^*, does consistency help? What criterion would help?
4. Tabulate the efficiency of p^* relative to $\widehat{p}$, for example when $n = 10$ and $p = 0.0, 0.1, 0.2, \ldots, 0.9, 1.0$.
5. State some possible circumstances when you might prefer to use p^* instead of $\widehat{p}$ to estimate p.

Problem 7

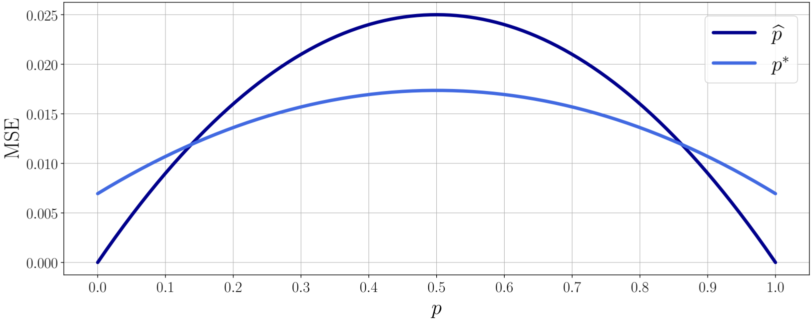

Figure: MSE of estimators $\widehat{p}$ and p^* for $n = 10$.

That's all Folks