Applied Statistics

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

Lecture Summary

- Reminder: MLE estimators.
- MLE estimators for the mean and variance of a normal distribution. (Chapter 7.5)
- Sampling distributions of estimators: definition, why do we care. (Chapter 8.1)
- Sampling distributions of estimators for the normal distribution. (Chapter 8.2, 8.3)
- The $Gamma$ and χ^2 distribution. (Chapter 5.7, 8.2)

Maximum Likelihood Estimation

Steps to identify an MLE estimate

- Assume $X_1, \ldots, X_n$ are independent and follow a distribution with pf $f(x|\theta)$.
- Compute the likelihood $LL(x_1, \ldots, x_n|\theta)$ (function of θ).
- Find a θ such that

$$\frac{dLL(\theta)}{d\theta} = 0, \quad \frac{d^2LL(\theta)}{d\theta^2} < 0$$

$\hat{\theta}$ is a function of the random sample.

Sampling Distributions

- Suppose $\mathbf{X} = (X_1, \dots, X_n)$ is a random sample from $f(x|\theta)$.
- A function $r(X_1, \dots, X_n)$ is a statistic.
- A sampling distribution: the distribution of a statistic (given θ).
- Can use the sampling distributions to compare different estimators and to determine the sample size we need.
- Used to get confidence intervals and to do hypothesis testing.
- Leads to definitions of new distributions, e.g. χ^2_m and t_m.

Sampling Distribution of a Statistic

Example:

Suppose we want to use a statistic $T = r(X_1, \ldots, X_n)$ as an estimate of a parameter θ. To be able to calculate things like

$$P(|T - \theta| < 0.05)$$

we need to know the distribution of T.

Back to Maximum Likelihood Estimation

Example: M.L.E. of normal

- Assume you observe the heights of n students.
- Let X_i be the height of the i-th student you picked.
- $X_i \sim Norm(\mu, \sigma^2)$ (known mean, unknown variance).
- We get data $x_1, \ldots, x_n$.
- Find the MLE $\hat{\sigma_0^2}$ of σ^2 (as a function of the data).

Sampling Distributions of the MLE estimators for normal

- Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$.
- If you know σ^2, $\overline{X}_n$ is the MLE for μ.
- If you know μ, $\hat{\sigma_0^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ is the MLE for σ^2.

Reminder

- Sum of normals is normal, $\overline{X}_n \sim N(\mu, \sigma^2/n)$.
- $\hat{\sigma_0^2}$?

The Gamma Distribution

Can model random variables that are known to be positive.

- The Gamma function: $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, \quad a > 0.$
- $\Gamma(n) = n!$
- $\Gamma(a) = (a-1)\Gamma(a-1)$

Gamma distributions

A continuous r.v. X has the Gamma distribution with parameters a and β if it has the pdf

$$f(x|a,\beta) = \frac{\beta^a}{\Gamma(a)} x^{a-1} e^{-\beta x}$$

Parameter space: $a > 0, \beta > 0$.

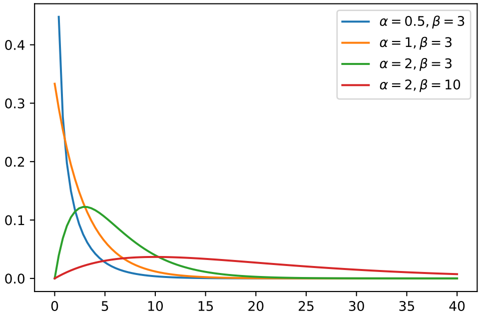
$$E(X) = \frac{a}{\beta}, \qquad Var(X) = \frac{a}{\beta^2}$$

The Gamma Distributions

- Reminder: Exponential distributions $f(x|\beta) = \beta e^{-\beta x}, x > 0$.
- Special case of the Gamma distributions with $a = 1$.

Theorem (5.7.7)

If $X_1, \ldots, X_k$ are independent random variables and $X_i \sim Gamma(a_i, \beta)$ then $X_1 + \cdots + X_k \sim Gamma(\sum_{i=1}^{N} a_i, \beta)$.



The χ^2 distributions

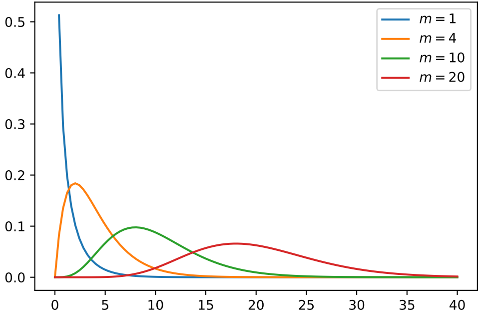
Definition

The χ^2 distribution with m degrees of freedom is the $Gamma(a = m/2, \beta = 1/2)$. The pdf is

$$f(x|m) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{m/2-1} e^{-x/2}$$

If $X \sim \chi^2_m$ then

- $E(X) = m$
- $Var(X) = 2m$



Properties of the χ^2 distributions

Theorem (8.2.1)

Let $X_1, \ldots, X_n$ *be independent random variables and* $X_i \sim \chi^2_{m_i}$. *Then*

$$X_1 + \cdots + X_n \sim \chi^2_m$$

where $m = m_1 + \cdots + m_n$.

- follows from Theorem 5.7.7.

Theorem (8.2.2)

If $X \sim \mathcal{N}(0,1)$, *then* $X^2 \sim \chi^2_1$

Properties of the χ^2 distributions

Corollary (8.2.1)

If the random variables $X_1, \ldots, X_n$ i.i.d., $X_i \sim \mathcal{N}(0,1)$ then

$$X_1^2 + \cdots + X_n^2 \sim \chi_n^2$$

The χ_m^2 distribution is a sampling distribution related to the sample variance of a normal distribution:

- If $X_1, \ldots, X_n$ are i.i.d, $X_i \sim \mathcal{N}(\mu, \sigma^2)$, where μ is known and the MLE of σ^2 is

$$\hat{\sigma_0^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

- Then

$$\frac{n\hat{\sigma_0^2}}{\sigma^2} \sim \chi_n^2$$

Back to Maximum Likelihood Estimation

Example: M.L.E. of normal

- Assume you observe the heights of n students.
- Let X_i be the height of the i-th student you picked.
- $X_i \sim Norm(\mu, \sigma^2)$ (unknown mean, unknown variance).
- We get data $x_1, \ldots, x_n$.
- Find the MLE of $(\theta) = (\mu, \sigma^2)$ (as a function of the data).

Sampling Distributions of the normal sample mean and variance

- Let $X_1, \ldots, X_n$ be a random sample from a $\mathcal{N}(\mu, \sigma^2)$ with unknown μ, σ^2.
- The sample mean and the sample variance are defined as

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

- They are the MLEs for μ and σ^2 in this setting.

Theorem (8.3.1)

Let $X_1, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$. Then $\bar{X}_n$ and S_n are independent random variables and $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $\frac{nS_n}{\sigma^2} \sim \chi^2_{n-1}$.

Sampling Distribution of the sample mean and variance

Theorem (8.3.1)

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Then $\overline{X}_n$ and S_n are independent random variables and

$$\overline{X_n} \sim \mathcal{N}(\mu, \sigma^2/n),$$

$$\frac{n}{\sigma^2} S_n = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 \sim \chi^2_{n-1}$$

- Replacing μ by $\overline{X}_n$ results in one less degree of freedom.
- $\overline{X}_n$ and S_n are functions of the same random variables, but they are independent (this only happens when the random sample is drawn from a normal distribution).

Recap

- Today we discussed samples of normal distributions:
- MLE estimators of μ and σ^2 in different settings.
- Sampling distributions of these estimators.
- Sample mean and sample variance are independent! (see an example in section 8.3 of your book)

Question: Bounding errors in estimates

- Suppose that $X_1, \ldots, X_m$ form a random sample from the normal distribution with mean μ and variance σ^2.
- Assuming that the sample size n is 16, determine the values of the following probabilities:

$$P(0.5\sigma^2 \leq \hat{\sigma_0^2} \leq 2\sigma^2)$$

$$P(0.5\sigma^2 \leq S_n \leq 2\sigma^2)$$

- Can you find the smallest valueof n such that

$$P(0.5\sigma^2 \leq S_n \leq 2\sigma^2) \geq 0.9?$$

Question: Sampling Distribution of an estimator

- Assume your population consists of 3 numbers: 1, 2, and 3.
- You draw a random sample of two numbers x_1, x_2 (with replacement) and take the sample mean $\overline{X}_n = \frac{1}{2}\sum_{i=1}^{2} x_i$
- Find the sampling distribution of $\overline{X}_n$.