Chapter 8: Sampling distributions of estimators

Sections

- 8.1 Sampling distribution of a statistic
- 8.2 The Chi-square distributions
- 8.3 Joint Distribution of the sample mean and sample variance
  - Skip: p. 476 - 478
- 8.4 The t distributions
  - Skip: derivation of the pdf, p. 483 - 484
- 8.5 Confidence intervals
- 8.6 Bayesian Analysis of Samples from a Normal Distribution
- 8.7 Unbiased Estimators
- 8.8 Fisher Information

Sampling distribution

- Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from $f(x|\theta)$
- 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 distributions

Sampling distribution

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

A familiar example:

- Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, \sigma^2)$ where σ^2 is known
- The sample mean $T = \overline{X}_n$ is a statistic and $N(\theta, \sigma^2/n)$ is the *sampling distribution* of T

The Chi-square distributions

Def: Chi-square distributions

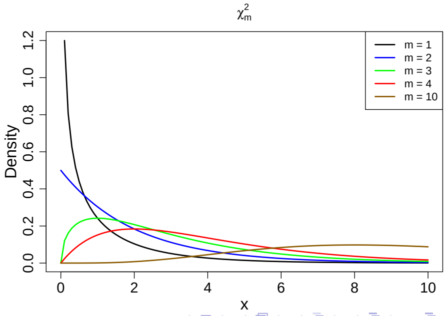
The χ^2_m *distribution with m degrees of freedom (df)* is the $\mathrm{Gamma}(\alpha = m/2, \beta = 1/2)$. The pdf is

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

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

- $E(X) = m$ and
- $\mathrm{Var}(X) = 2m$ and
- $\psi(t) = \left(\frac{1}{1-2t}\right)^{m/2}$

The χ^2_m distribution is tabulated at the end of the book



Properties of the χ^2_m distributions

And connections to the normal distributions

Theorem 8.2.1: A sum of chi-squares is a chi-square

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 directly from the fact that a sum of $\mathrm{Gamma}(\alpha_i, \beta)$ random variables (same β) is a $\mathrm{Gamma}(\sum_{i=1}^n \alpha_i, \beta)$

Theorem 8.2.2: Square of a standard normal is a chi-square

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

Properties of the χ^2_m distributions

And connections to the normal distributions

Corollary 8.2.1

If the random variables $X_1, \ldots, X_n$ i.i.d. $N(0, 1)$ then

$$X_1^2, \ldots, X_n^2 \sim \chi^2_n$$

When do we use the χ^2_m distribution?

- If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$ where μ is known then the MLE of σ^2 is

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

- We can easily show that

$$\frac{n\,\widehat{\sigma_0^2}}{\sigma^2} \sim \chi^2_n$$

- What then is the distribution of $\widehat{\sigma_0^2}$?

Sample mean and sample variance

- Let $X_1, \ldots, X_n$ be a random sample
- The *sample mean* and the *sample variance* are defined as

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

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

$$\overline{X}_n \sim N(\mu, \sigma^2/n) \quad \text{and}$$

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

Sample mean and sample variance

About Theorem 8.3.1:

- $\overline{X}_n$ and S_n are the MLE's of μ and σ^2
- $\overline{X}_n \sim N(\mu, \sigma^2/n)$ was already known
- We knew that $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2_n$. The effect of replacing μ with $\overline{X}_n$ is that the degrees of freedom go from n to $n-1$
- Even though $\overline{X}_n$ and S_n are functions of the same random variables they are independent.

Example

Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ and let

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

a) Assuming $n = 16$ determine

$$P\left(\frac{1}{2}\sigma^2 \leq \hat{\sigma}^2 \leq 2\sigma^2\right)$$

b) Determine the smalles value of n so that

$$P\left(\frac{1}{2}\sigma^2 \leq \hat{\sigma}^2 \leq 2\sigma^2\right) \geq 0.9$$

The student's *t* distributions

The *t* distributions

Let $Y \sim \chi^2_m$ and $Z \sim N(0, 1)$ be independent. Then the distribution of

$$X = \frac{Z}{\left(\frac{Y}{m}\right)^{1/2}}$$

is called the *t distribution with m degrees of freedom, or* t_m

- You can see where this is going: We want the distribution of $\frac{X-\mu}{\sigma}$ where the σ is replaced by the sample standard deviation.
- Introduced by W. S. Gosset, who wrote under the alias "Student"
- The legend:
    - Gosset derived the t_m distribution while working for the Guinness Brewery in Dublin. I fear of competition he was forbidden to publish his analysis of brewery data and hence he wrote under the pseudonym Student.

The student's t distributions

- The pdf of the t_m distribution is

$$\frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-(m+1)/2} \qquad -\infty < x < \infty$$

- Tabulated in the back of the textbook

If $X \sim t_m$ then

- $E(X) = 0$ if $m > 1$, does not exist otherwise
- $\mathrm{Var}(X) = \frac{m}{m-2}$ if $m > 2$, does not exist otherwise

Connection to the normal random variables

Theorem 8.4.2

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ and let $\overline{X}_n$ be the sample mean and let

$$\sigma' = \left[\frac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{n-1}\right]^{1/2}$$

Then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma'} \sim t_{n-1}$$

Note that σ' is not the MLE of σ. In fact

$$\sigma' = \left(\frac{n}{n-1}\right)^{1/2} \hat{\sigma}$$

For large n, σ' and $\hat{\sigma}$ are close.

Connection to the normal and Cauchy

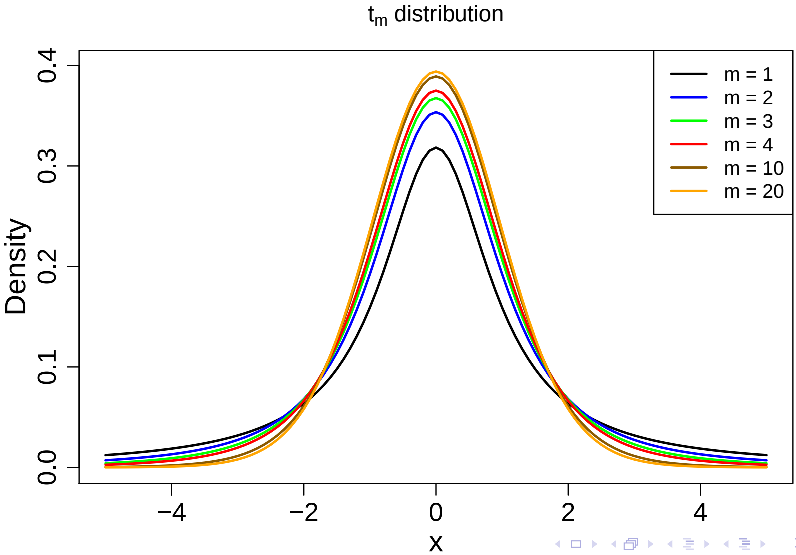
- As $m \to \infty$ the t_m approaches $N(0,1)$
- $t_1 = \text{Cauchy}$:

$$f(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma\left(\frac{m}{2}\right)}\left(1 + \frac{x^2}{m}\right)^{-(m+1)/2}$$

if $m = 1$:

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

The t_m distributions



The t_m distributions