Statistical Methods MATH2715 info

Teaching material is all online!

- On Minerva http://minerva.leeds.ac.uk
- On GitHub
  https://github.com/luisacutillo78/Statistical-Methods-Lect

R code submission

- No technichal issue - please submit your SURNAMEstudentid.R [or .Rmd as required] file in the assignment folder.
- **Please print a copy of your notebook and put it into your marker collection box**.

Resources

- Mathematical Statistics and Data Analysis - 3rd ed. (by J. A. Rice);
- http://www1.maths.leeds.ac.uk/statistics/R/Rintro.pdf;
- https://www.datacamp.com/courses/free-introduction-to-r.

Where We've Been, Where We're Going

In the previous Lecture

- Univariate and Multivariate Change of Variables
- Weak Law of Large Numbers: Interpretation and discussion
- Convergence in probability
- Convergence in distribution
- Central limit theorem
- Jupyter Notebooks implementation of the CLT and Chebyshev inequality

Today

- Random Samples from Normal Distributions
- Socrative Quiz

Random Samples from Normal Distributions

Why is it important

- The CLT central limit theorem suggests that for large samples a normal approximation may be appropriate!
- Lots of known properties!

Some Results: Linear combination of Normal iid

Theorem A

If $X_1, X_2, \ldots, X_n$ are independent random variables with $X_j \sim N(\mu_j, \sigma_j^2)$, then $Y = \sum_{j=1}^n \alpha_j X_j$ is also normally distributed with mean $\sum_{j=1}^n \alpha_j \mu_j$ and variance $\sum_{j=1}^n \alpha_j^2 \sigma_j^2$

Proof on notes.

Definition

A chi-square distribution with r degrees of freedom is a gamma distribution: $\chi^2(r)$ is the same as $\Gamma\left(\frac{r}{2}, \frac{1}{2}\right)$.

Note that the moment generating function is

$$M(t) = \left(\frac{1}{1-2t}\right)^{\frac{r}{2}}.$$

$\square$

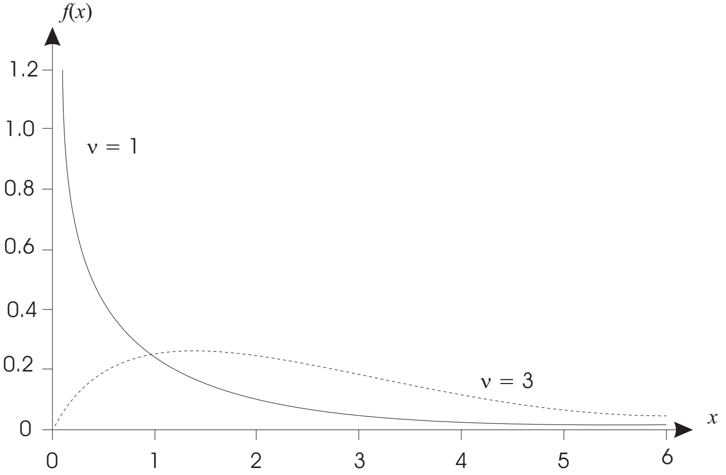
The density function $\chi^2(1)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{1}{2}x}, \quad x > 0.$$

Using the above moment generating function, we see that $E(X) = M'_X(0) = 1$, $V(X) = 2$.

Chi-square distribution

You can see from the graphs below that χ^2-distributions have a distinctively skewed shape.



The solid line is $\chi^2(1)$ pdf and the dotted line is $\chi^2(3)$ pdf.

Lemma A

If $X \sim N(0,1)$, then $Y = X^2 \sim \chi^2(1)$.
□

Proof.

Let $Y = X^2$. Then

$$F_Y(y) = P(Y \le y) = P\left(X^2 \le y\right) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X($$

so that

$$f_Y(y) = F_Y'(y) = \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}, \quad y > 0.$$

■

□

Sum of $\chi^2(1)$ rv

Theorem B

If $Y_1, Y_2, \ldots, Y_r$ are an independent random sample, each with a $\chi^2(1)$ distribution then

$$\sum_{j=1}^{r} Y_j \sim \chi^2(r).$$

Proof.

on notes. $\square$

Corollary

If $Y_1 \sim \chi^2(r)$ and $Y_2 \sim \chi^2(s)$, and are independent, then

$$Y_1 + Y_2 \sim \chi^2(r+s).$$

Independence of $\overline{X}$ and S^2 for normal samples

Recall that:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

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

Key results:

- Independence of $\overline{X}$ and S^2 for normal samples
- their relationship to the mean and variance parameters of a normal distribution.

Independence of $\overline{X}$ and S^2 for normal samples

Theorem C

If $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with normal distribution $N\left(\mu, \sigma^2\right)$, then the vector of rv $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$ and $\overline{X}$ are independent.

Independence of $\overline{X}$ and S^2 for normal samples

Theorem D

If $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with normal distribution $N\left(\mu, \sigma^2\right)$, then $\overline{X}$ and S^2 are independent with distributions

(i) $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$;

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

Proof.

S^2 is function of $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$. According to Theorem C, the vector of rv $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$ and $\overline{X}$ are independent. It follows that S^2 and $\overline{X}$ are independent. $\square$

Theorem D

Proof.

Note that (i) follows from Theorem A. Moreover it can be shown (see notes) that:

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

hence it follows that:

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

$\square$

t-distribution

Definition

If $U \sim N(0,1)$ and $V \sim \chi^2(r)$ are independent, then

$$T = \frac{U}{\sqrt{V/r}} \sim t(r)$$

has a t-distribution with r degrees of freedom. This defines the distribution $t(r)$.

t-distribution

The following Figure shows a graph of $t\,(7)$ compared with a standard $N\,(0,1)$ distribution. You can see that it is very similar, but has fatter tails.

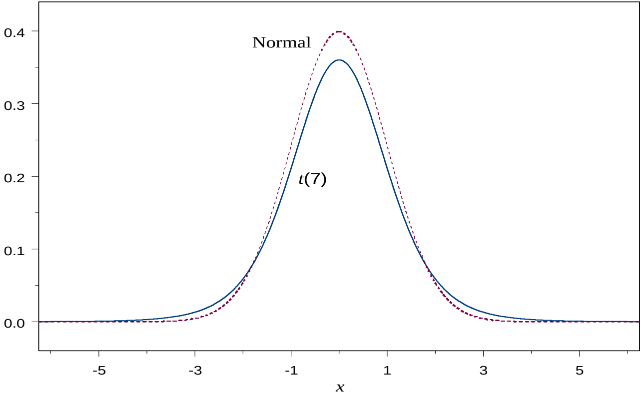


Figure 2 Normal and $t(7)$ distributions

Properties of the t-distribution

(i) The pdf of the $t(r)$ distribution is

$$f(t) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{(\pi r)}\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}}, \ t \in \mathbb{R}.$$

(ii) If $r = 1$ then $t(1)$ is a Cauchy distribution without finite mean or variance.

(iii) As $r \to \infty$ then $t(r) \to \mathrm{N}(0, 1)$.

Properties of the t-distribution

Theorem

If $X_1, X_2, \ldots, X_n$ are a normal random sample then

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t(n-1).$$

Proof.

$\overline{X}$ and S^2 are independent and

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0,1) \ , \ \frac{(n-1)\, S^2}{\sigma^2} \sim \chi^2(n-1).$$

Using the definition of the t-distribution, the result follows with the unknown σ cancelling. $\square$

Join Socrative, Sudent login, ROOM NAME: CUTILLO01

- `https://www.socrative.com/`
- Socrative Student on the App Store - iTunes - Apple
- Socrative Student - Apps on Google Play