Small-Sample Statistics

Student's T Distribution

Bruce M. Boghosian

Tufts UNIVERSITY | School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Using the sample variance for estimation

- Recall difference between *variance* and *sample variance*
  - $\sigma_Y^2 = \frac{1}{n}\sum_{j=1}^{n}\left(Y_j - \overline{Y}\right)^2$
  - $S_Y^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_j - \overline{Y}\right)^2$
  - where $\overline{Y} := \frac{1}{n}\sum_{k=1}^{n} Y_k$
- By the CLT, $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is normally distributed.
- Question: Is $T = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$ also normally distributed?
- Answer: For very large n, there is little difference in distributions of Z and T.

Large samples and small samples

Bruce M. Boghosian

- For many years it was believed T was also normally distributed for small n.
- William Sealy Gossett (1876-1937) first to realize it is not.
- Quality assurance statistician at Guinness brewery, Dublin
  - Barley, etc. in small batches (small n) from small farms.
  - σ generally unknown and had to be inferred from the data
  - Pdf of $T = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$ decayed slower than normal pdf.
  - Pdf was still bell-shaped, but the tails were "thicker".

Review of methodology

- Prior to this module:
  - Review Γ distributions and sums of Γ-distributed r.v.s
  - Understand χ^2 distribution as special case of Γ distribution
  - Sums of squares of iid normal r.v.s, $\sum_{j=1}^n Z_j^2$, are χ^2 dist.
  - Show $\overline{Y}$ and S_Y^2 are independent
  - Show $\frac{(n-1)S^2}{\sigma^2}$ is chi square distributed
  - Pdf of ratio of two iid chi square r.v.s – F distribution
- In this module:
  - Show that $T^2 = \left(\frac{\overline{Y}-\mu}{S/\sqrt{n}}\right)^2$ is F distributed
  - Use the above to derive the T distribution pdf f_T

Definition and symmetry of the T distribution

Bruce M. Boghosian

- **Def.:** Let Z be an r.v. sampled from $N(0,1)$, and let U be a chi squared r.v., independent of Z with n degrees of freedom. The Student T ratio with n df is denoted T_n where $T_n = \frac{Z}{\sqrt{U/n}}$.
- **Lem.:** The pdf $f_{T_n}(t)$ is a symmetric function. That is $\forall t: \; f_{T_n}(-t) = f_{T_n}(t)$.
- **Pf.:** We know f_Z is symmetric. Let $V = \sqrt{U/n}$ so

$$f_{T_n}(t) = \int_0^\infty dv \; f_V(v) f_Z(tv) = \int_0^\infty dv \; f_V(v) f_Z(-tv) = f_{T_n}(-t) \quad \square$$

The pdf of the T distribution I

- **Thm.:** The pdf for a Student T random variable with n degrees of freedom is given by

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{(n+1)/2}} \quad \text{for } t \in \mathbb{R}.$$

- **Pf.:** Note that $T_n^2 = \frac{Z^2}{U/n}$ has an F distribution with 1 and n df. Hence,

$$f_{T_n^2}(t) = \frac{n^{n/2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{t^{-1/2}}{(n+t)^{(n+1)/2}}$$

The pdf of the T distribution II

Bruce M. Boghosian

- **Pf. (continued):** By the symmetry of T_n, we have

$$\begin{aligned} F_{T_n} = P(T_n \leq t) &= \frac{1}{2} + P(0 \leq T_n \leq t) \\ &= \frac{1}{2} + \frac{1}{2} P(-t \leq T_n \leq t) \\ &= \frac{1}{2} + \frac{1}{2} P(0 \leq T_n^2 \leq t^2) \\ &= \frac{1}{2} + \frac{1}{2} F_{T_n^2}(t^2) \end{aligned}$$

- Differentiating yields

$$f_{T_n}(t) = \frac{1}{2} f_{T_n^2}(t^2)(2t) = t\, f_{T_n^2}(t^2).$$

- **Pf. (continued):** Complete proof as follows

$$
\begin{aligned}
f_{T_n}(t) =& t\, f_{T_n^2}(t^2) \\
=& t\, \frac{n^{n/2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{(t^2)^{-1/2}}{(n+t^2)^{(n+1)/2}} \\
=& \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \qquad \square
\end{aligned}
$$

as was to be shown.

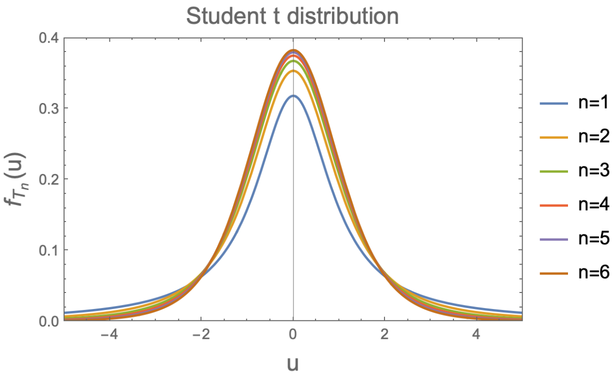
$\overline{Y}$ is Student T distributed

- **Thm.:** Let $Y_1, \ldots, Y_n$ be a random sample, each from $N(\mu, \sigma^2)$. Then, $T_{n-1} = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$ has a Student T distribution with $n-1$ df.
- **Pf.:** Note that we can write

$$\frac{\overline{Y}-\mu}{S/\sqrt{n}} = \frac{\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}$$

- Numerator is in $N(0,1)$ and denominator has a chi squared distribution with $n-1$ df.
- Numerator and denominator are independent r.v.s.
- The statement of the theorem follows instantly. $\square$

Plotting the T distribution



Tables in the appendices

- You are now in a position to understand yet another table in the back of the book.
- Table A.1 tabulates Z distributions for various α.
- Table A.3 tabulates χ^2 distributions for various α and n df.
- Table A.4 tabulates F distributions for various α and m and n df.

Summary

- We have completed the derivation of the student T distribution.
- It can be used for sampling small-sample, normally distributed data.
- It is tabulated and it is possible to use it for interval estimation, hypothesis testing, etc.