5.2 Useful classical facts

Theorem if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

Let $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$

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

Then

1. $\bar{X}$ and S^2 are independent Random variables
2. $\bar{X} \sim N(\mu, \sigma^2/n)$
3. $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$ [chi square distribution with $(n-1)$ degree of freedom (d.f.)]

Recall from Section 3-2a.

χ^2_p denotes chi square distribution with p d.f.

the pdf:

$$f(x) = \frac{1}{\Gamma(p/2)\, 2^{p/2}} x^{p/2-1} e^{-x/2}, \quad x > 0$$

$$\chi^2_p \equiv \text{Gamma}(p/2, 2).$$

Proof (Sketch) Suffice to prove for $\sigma = 1$.

1. Note

$$(n-1)S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$= \underbrace{(X_1 - \bar{X})^2}_{} + \sum_{i=2}^{n} \underbrace{(X_i - \bar{X})^2}_{Y_i}$$

$$= \left(\sum_{i=2}^{n} Y_i\right)^2 + \sum_{i=2}^{n} Y_i^2$$

Use change of variable formula for the mapping $(X_1, \ldots, X_n) \mapsto (Y_1, Y_2, \ldots, Y_n)$, where $\begin{cases} Y_1 = \bar{X} \\ Y_i = X_i - \bar{X}, \ i \geq 2 \end{cases}$

to find that

$$f_{Y_1 \ldots Y_n}(y_1, \ldots, y_n) = \left(\frac{n}{2\pi}\right)^{1/2} e^{-n y_1^2/2} \frac{n^{1/2}}{(2\pi)^{(n-1)/2}} e^{-\frac{1}{2}\left(\left(\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2\right)}$$

which factorizes.

Hence $Y_1 \perp\!\!\!\perp (Y_2, \ldots, Y_n)$

So $(n-1)S^2 \perp\!\!\!\perp Y_1 = \bar{X}$.

2. Done in previous section

3. Let $\overline{X}_k = \frac{1}{k}(X_1 + \dots + X_k)$

$$S_k^2 = \frac{1}{k-1} \sum_{i=1}^{k} (X_i - \overline{X}_k)^2$$

Proof by induction that $(k-1) S_k^2 \sim \chi^2_{k-1} \quad \forall k \geq 2.$

- For $k=2$, $S_2^2 = (X_1 - \overline{X}_2)^2 + (X_2 - \overline{X}_2)^2$

$$= \frac{1}{2}(X_1 - X_2)^2$$

$X_1 - X_2 \sim N(0,2) \Rightarrow \left(\frac{1}{\sqrt{2}}(X_1 - X_2)\right)^2 \sim N(0,1)^2$ which is χ^2_1

- Suppose now $(k-2) S_{k-1}^2 \sim \chi^2_{k-1}$

  We'll be done by proving that $(k-1) S_k^2 \sim \chi^2_k$.

The key is the recursive relation, which can be verified

$$\boxed{(k-1) S_k^2 = (k-2) S_{k-1}^2 + \left(\frac{k-1}{k}\right) (X_k - \overline{X}_{k-1})^2}$$

Now we proceed by the following sequence of arguments

* $X_k - \overline{X}_{k-1} \sim N\left(0, \frac{k}{k-1}\right)$ for $k \geq 2$

$$\Rightarrow \sqrt{\frac{k-1}{k}} (X_k - \overline{X}_{k-1}) \sim N(0,1)$$

$$\Rightarrow \frac{k-1}{k} (X_k - \overline{X}_{k-1})^2 \sim \chi^2_1$$ (Chi-square df 1)

See Sec 2-1.

* Check $(X_k - \overline{X_{k-1}}) \perp\!\!\!\perp S_{k-1}$ (by checking joint pdf)

* By inductive hypothesis, $(k-2)\, S_{k-1}^2 \sim \chi^2_{k-1}$

* Now
$$\left.\begin{cases} \chi^2_{k-1} \equiv \text{Gamma}\left(\frac{k-1}{2}, 2\right) \\ \chi^2_1 \equiv \text{Gamma}\left(\frac{1}{2}, 2\right) \end{cases}\right\} \text{Sec 3.2a}$$

* Now sum of 2 indep Gamma variables $\left(\frac{k-1}{2}, 2\right)$ and $\left(\frac{1}{2}, 2\right)$ is another Gamma $\left(\frac{k}{2}, 2\right)$, (via MGF argument) which is χ^2_k.

Hence $(k-1)\, S_k^2 \sim \chi^2_k$ as we need to show. □

Other Statistics and their distributions

- if $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, μ unknown
  $\bar{X}$ is a statistic that tells us about μ.

  How so? we know $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

  this gives us a way to quantify the uncertainty about μ, using statistic $\bar{X}$ as an estimate.

- what if σ is unknown too

  we may want to consider $\dfrac{\bar{X} - \mu}{S/\sqrt{n}}$

  since S is a statistic that can be obtained from the sample.

  To be useful, we need to know the distribution of

$$T := \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Theorem $T \sim t_{n-1}$ (Student's t distribution with $n-1$ degree of freedom)

t_p has the pdf

$$f_T(t) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)} \frac{1}{(p\pi)^{1/2}} \frac{1}{\left(1+\frac{t^2}{p}\right)^{(p+1)/2}} .$$

$-\infty < t < \infty$.

Remark

- T is the Ratio of two indep. Random variables, as $\bar{X}-\mu \sim$ Normal, $\sqrt{(n-1)S^2}$ square Root of χ^2_{n-1}
- if $n=2$ then $T \sim$ Ratio of 2 indep normal variables $\Rightarrow T \sim$ Cauchy.
- Proof is by change of variable formula.

For more, see Section 5.3 in the text book.