Applications of statistical inference. Summary of necessary theoretical results, chapter 4.1

Grethe Hystad

October 7, 2012

- So far we have looked at point estimators of parameters through maximum likelihood estimation.
- In chapter 4.2 and 4.3, we will look at interval estimation of the parameters through confidence intervals.
- In chapter 4.3-4.5, we will also create statistical hypothesis about parameters using the distribution of the estimators.

Theorem

Theorem 1.
Let X_1, X_2,..., X_n be independent and normal random variables with means μ_1, μ_2,..., μ_n and variances σ_1^2, σ_2^2..., σ_n^2 respectively. Then the random variable

$$Y = \sum_{i=1}^{n} a_i X_i$$

has a normal distribution with mean

$$\mu_Y = \sum_{i=1}^{n} a_i \mu_i \quad \text{and variance} \quad \sigma_Y = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Theorem

Theorem 2.

Let X_1, X_2,..., X_n be n independent and chi-square random variables with $r_1, r_2, ..., r_n$ degrees of freedom, respectively. Then the random variable

$$Y = \sum_{i=1}^{n} X_i$$

has a chi-square distribution with $r_1 + r_1 + \cdots + r_n$ degrees of freedom.

Example

The ounces of fill per box of filling cereal boxes are normally distributed. Suppose company A has machines that fill the boxes with a mean of 14.5 ounces of fill per box and standard deviation of 1.3 ounces of fill per box. Suppose Company B fills the boxes with a mean of 14 ounces of fill per box and standard deviation of 1.0 ounces of fill per box.
Let X_1 and X_2 be the ounces of fill per box for company A and company B, respectively. Define $Y = X_1 - X_2$.
Find the mean and variance of Y.

Solution

Y is a normal random variable. We have
$E(Y) = E(X_1) - E(X_2) = 14.5 - 14 = 0.5$ *and*
$Var(Y) = (1^2)Var(X_1) + (-1)^2 Var(X_2) = (1.3)^2 + (1.0)^2 = 2.69$

Example

Let X_1, X_2, X_3 be three independent chi-square random variables with $r_1 = 1$, $r_2 = 3$ and $r_3 = 7$ degrees of freedom. Define $Y = X_1 + X_2 + X_3$. Then Y is $\chi(11)$ and

$$P(Y < 19.68) = 0.950$$

by table IV.

Theorem

Theorem 3.

Let X_1, X_2, ..., X_n be i.i.d. normal random variables $N(\mu, \sigma^2)$. Then

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

are independent random variables such that

- $\overline{X}$ *is* $N(\mu, \frac{\sigma^2}{n})$ *and*
- $\frac{(n-1)S^2}{\sigma^2}$ *is* $\chi^2(n-1)$, *i.e.*

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n-1} Z_i^2,$$

where $Z_1,\ldots,Z_{n-1}$ are independent standard normal random variables, $N(0,1)$.

- If X_1, X_2,...,X_n are i.i.d. from $N(\mu, \sigma^2)$. Then $Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2}$ is $\chi^2(1)$ for $i = 1, 2, ..., n$.

- If X_1, X_2,...,X_n are i.i.d. from $N(\mu, \sigma^2)$. Then $Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2}$ is $\chi^2(1)$ for $i = 1, 2, ..., n$.
- From Theorem 2, we have $Z_1^2 + \cdots + Z_n^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2}$ is $\chi^2(n)$ since each Z_i are independent $X^2(1)$ for $i = 1, ..., n$.

- If X_1, X_2,...,X_n are i.i.d. from $N(\mu, \sigma^2)$. Then $Z_i^2 = \frac{(X_i-\mu)^2}{\sigma^2}$ is $\chi^2(1)$ for $i = 1, 2, ..., n$.
- From Theorem 2, we have $Z_1^2 + \cdots + Z_n^2 = \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2}$ is $\chi^2(n)$ since each Z_i are independent $X^2(1)$ for $i = 1, ..., n$.
- From Theorem 3, replacing μ by its estimator $\overline{X}$, we have

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

which is $\chi^2(n-1)$.

- If X_1, X_2,...,X_n are i.i.d. from $N(\mu, \sigma^2)$. Then $Z_i^2 = \frac{(X_i-\mu)^2}{\sigma^2}$ is $\chi^2(1)$ for $i = 1, 2, ..., n$.
- From Theorem 2, we have $Z_1^2 + \cdots + Z_n^2 = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}$ is $\chi^2(n)$ since each Z_i are independent $X^2(1)$ for $i = 1, ..., n$.
- From Theorem 3, replacing μ by its estimator $\overline{X}$, we have

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

  which is $\chi^2(n-1)$.
- We have lost one degree of freedom by replacing a parameter in a chi-square variable by its estimator.

- If X_1, X_2,...,X_n are i.i.d. from $N(\mu, \sigma^2)$. Then $Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2}$ is $\chi^2(1)$ for $i = 1, 2, ..., n$.
- From Theorem 2, we have $Z_1^2 + \cdots + Z_n^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$ is $\chi^2(n)$ since each Z_i are independent $X^2(1)$ for $i = 1, ..., n$.
- From Theorem 3, replacing μ by its estimator $\overline{X}$, we have

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

  which is $\chi^2(n-1)$.
- We have lost one degree of freedom by replacing a parameter in a chi-square variable by its estimator.
- In general: If we replace p parameters in a $\chi^2(r)$ $(p < r)$ random variable by estimators, then the resulting chi-square variable has $r - p$ degrees of freedom.

Student's t distribution with r degrees of freedom

Define the random variable

$$T = \frac{Z}{\sqrt{U/r}},$$

where

- Z is $N(0, 1)$ and
- U is $\chi^2(r)$ and
- Z and U are independent.

The distribution of T is called the **Student's t distribution with** r **degrees of freedom**.

Student's t distribution with r degrees of freedom

Example

If $r = 5$ we have $P(T \leq 2.015) = 0.95$ from table *VI* in appendix B.

In general, we find $t_\alpha(r)$ such that $P(T > t_\alpha(r)) = \alpha$

Student's t distribution with r degrees of freedom

From Theorem 3,

- If $X_1, \dots, X_n$ are iid $N(\mu, \sigma^2)$, then $\overline{X}$ is $N(\mu, \frac{\sigma^2}{n})$ so $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is $N(0,1)$ and
- $U = \frac{(n-1)S^2}{\sigma^2}$ is $\chi^2(n-1)$ and
- $\overline{X}$ and S^2 are independent.

Hence

$$T = \frac{Z}{\sqrt{\frac{U}{(n-1)}}} = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$$

has a Student's t distribution with $n-1$ degrees of freedom.

Student's t distribution with r degrees of freedom

The distribution function of T is given by

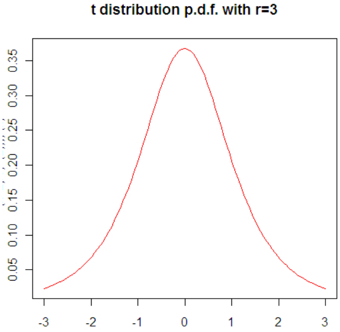
$$F(t) = P(T \leq t) = P(Z \leq t\sqrt{U/r})$$

The p.d.f of T is $f(t) = F'(t)$ which is

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}} \quad \text{for} \quad -\infty < t < \infty.$$

Then mean is defined when $r > 1$ and the variance when $r > 2$.

Student's t distribution with r degrees of freedom



The tails of the t distribution are heavier than those of the Normal distribution.

As $n \to \infty$, the Student's t distribution converges to the standard Normal distribution.

- For large sample sizes $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ is approximately $N(0,1)$ by the Central limit theorem.
- Often the parameter σ is unknown and we replace σ by its estimate, the sample standard deviation s.
- When $\overline{X} = \overline{x}$, we then have the statistics, $T = t$, where $t = \frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}}$.
- If the sample size is large, the distribution of T is approximately standard normal.
- If the sample size is small and the underlying distribution is not normal or we don't have information about it, then we don't know much about the distribution of T.

Student's t distribution with r degrees of freedom

- If the sample size is less than $n = 15$, use the student's t distribution if the data are close to Normal. If the data are skewed or if there are outliers in the data, the t-distribution is not a good approximation.
- If the sample size is at least $n = 15$, the t-distribution is a good approximation except if there are strong skewness or outliers in the data.
- If the sample size is large, $n \geq 40$, the t-distribution is a good approximation even for clearly skewed distributions.

Student's t distribution with r degrees of freedom

R-code:
$qt(p, df, lower.tail = TRUE)$ computes the $(100p)th$ percentile, where df is the number of degrees of freedom.

$> qt(0.975, 3, lower.tail = TRUE)$
[1] 3.182446

Student's t distribution with r degrees of freedom

$pt(q, df, lower.tail = TRUE)$ computes the probability, $p = P(T \leq q)$.

$> pt(3.182446, 3, lower.tail = TRUE)$
[1] 0.975

The command lower.tail can be omitted (The default setting is lower.tail=TRUE).
To calculate $P(T > q)$, set lower.tail=FALSE as in $pt(q, df, lower.tail = FALSE)$

$> pt(3.182446, 3, lower.tail = FALSE)$
[1] 0.02500001

$> qt(0.025, 3, lower.tail = FALSE)$
[1] 3.182446

Fisher's F distribution

Define

$$F = \frac{U_1/r_1}{U_2/r_2},$$

where

- U_i is $\chi^2(r_i)$ for $i = 1, 2$ and
- U_1 and U_2 are independent.

The distribution of F is called the Fisher's F distribution with r_1 and r_2 degrees of freedom. We denote the distribution by $F(r_1, r_2)$.

Fisher's F distribution

Example

Let $r_1 = 2$ and $r_2 = 7$.
Then $P(F \leq 4.74) = 0.95$ and
$P(F > 9.55) = 0.01$ from table VII in appendix B.

We can find F_α so that $P(F > F_\alpha) = \alpha$

Fisher's F distribution

Let W=F. The distribution function of W is given by

$$F(w) = P(W \leq w) = P(U_1 \leq [r_1/r_2]wU_2).$$

The p.d.f. of W is $f(w) = F'(w)$, where

$$f(w) = \frac{(r_1/r_2)^{r_1/2}\Gamma((r_1 + r_1)/2)w^{(r_1/2)-1}}{\Gamma(r_1/2)\Gamma(r_2/2)(1 + (r_1 w/r_2))^{(r_1+r_2)/2}}$$