Chi-squared and t-distributions

Numerical Statistics Fall, 2021

Jake Underland

2021-11-08

Contents

The Chi-squared Distribution

Let

$$Z_1, Z_2, \dots, Z_n \overset{i.i.d.}{\sim} N(0,1)$$

Then,

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$$

Expectation

First,

$$Var(Z) = E(Z^2) - E(Z)^2 = E(Z^2) - 0 = 1 \implies E(Z^2) = 1$$

Then,

$$\begin{aligned} E(Y) &= E(\sum_i^n Z_i^2) \\ &= \sum_i^n E(Z_i^2) \\ &= \sum_i^n 1 \\ &= n \end{aligned}$$

Variance

$$
\begin{aligned}
Var(Y) &= \sum_i Var(Z_i^2) \ldots Independence \\
Var(Z_i^2) &= E(Z^4) - E(z^2)^2 \\
&= E(Z^4) - 1^2 \\
\Gamma(z) &= \int_0^\infty x^{z-1} e^{-x} dx;\ \Gamma(z+1) = z\Gamma(z) \\
\Gamma(s)\Gamma(1-s) &= \frac{\pi}{\sin \pi s} \implies \Gamma(\frac{1}{2})\Gamma(1-\frac{1}{2}) = \frac{\pi}{\sin\frac{\pi}{2}} \implies \Gamma(\frac{1}{2})^2 = \pi \implies \Gamma(\frac{1}{2}) = \sqrt{\pi} \\
E(Z^4) &= \int_{-\infty}^\infty z^4 \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dz \\
&= \frac{2}{\sqrt{2\pi}} \int_0^\infty z^4 \exp(-\frac{z^2}{2}) dz \mid_{u-sub\ u=z^2/2} \\
&= \frac{2}{\sqrt{2\pi}} \cdot 2\sqrt{2} \int_0^\infty u^{\frac{3}{2}} e^{-u} du \\
&= \frac{4}{\sqrt{\pi}} \int_0^\infty u^{\frac{3}{2}} e^{-u} du \\
&= \frac{4}{\sqrt{\pi}} \Gamma(\frac{5}{2}) \\
&= \frac{4}{\sqrt{\pi}} \cdot \underbrace{\frac{3}{2} \cdot \frac{1}{2} \cdot \Gamma(\frac{1}{2})}_{\text{recursiveness of } \Gamma()} \\
&= \frac{4}{\sqrt{\pi}} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \Gamma(\frac{1}{2}) \\
&= \frac{4}{\sqrt{\pi}} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} \\
&= 3 \\
\implies Var(Y) &= \sum_i Var(Z_i^2) \\
&= \sum_i \big(E(Z^4) - 1\big) \\
&= \sum_i (3-1) \\
&= \sum_i 2 \\
&= 2n \ldots \square
\end{aligned}
$$

Sampling Distribution of Variances

Let

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

Then,

$$\sum_i \left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_i Z_i^2 \sim \chi_n^2$$

$$\boldsymbol{\iff \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2}$$

Now,

$$
\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu)^2 \\
&= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\bar{X})+n(\bar{X}-\mu)^2\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})+\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2
\end{aligned}
$$

Where

$$
\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)
$$

$$
\implies \left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi_1^2
$$

Thus, following from the independence of $\bar{X}$ and $\sum_{i=1}^{n}(X_i-\bar{X})^2$, proven in lectures for STAT 245, we have

$$
\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}_{\sim\chi_n^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})+\underbrace{\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2}_{\sim\chi_1^2}
$$

$$
\implies \boldsymbol{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}) \sim \chi_{n-1}^2}
$$

Additionally,

$$
\hat{S}^2 = \frac{1}{n-1}\sum_{i}^{n}(X_i-\bar{X})^2 \implies \frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi_{n-1}^2
$$

t-distribution

Definition

Let

$$
Y \sim \chi_n^2;\ Z \sim N(0,1);\ Y \perp\!\!\!\perp Z
$$

Then,

$$
\frac{\sqrt{n}Z}{\sqrt{Y}} \sim t_n
$$

Sampling with unknown population variance

Let

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

Then,

$$
\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)
$$

Let $\hat{S}=\sqrt{\hat{S}^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}$
Then, from the definition of the t distribution,

$$
\boldsymbol{T = \frac{\sqrt{n}(\bar{X}-\mu)}{\hat{S}} \sim t_{n-1}}
$$

Expectation

Let $T \sim t_m$:

$$E(T) = 0$$

Variance

$$Var(T) = \frac{m}{m-2} \ldots (m > 2)$$

Relation to Normal

From Chebyshev's inequality, we have $\forall \epsilon > 0$

$$P(|\hat{S}^2 - \sigma^2| > \epsilon) \leq \frac{Var(\hat{S}^2)}{\epsilon^2} = \frac{1}{\epsilon^2} \cdot \frac{2\sigma^4}{n-1} \to 0 \ \ (n \to \infty)$$

Where

$$\begin{aligned} Var\left(\frac{(n-1)\hat{S}^2}{\sigma^2}\right) &= 2(n-1) \\ \frac{(n-1)^2}{\sigma^4} Var(\hat{S}^2) &= 2(n-1) \\ \implies Var(\hat{S}^2) &= \frac{2\sigma^4}{(n-1)} \end{aligned}$$

This means that $\hat{S}^2 \xrightarrow{p} \sigma^2$ (consistent), and as $n \to \infty$,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\hat{S}} \to \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

Median

When the size of a data is odd ($n = 2m + 1$ where $m \in \mathbb{N}$), then the median of the data is

$$\arg\min_c f(c) = \sum_{i=1}^{n} |x_i - c|$$