ST5201: Basic Statistical Theory
Chapter 6: Distributions Derived from the Normal Distribution

CHOI Yunjin
stachoiy@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

24th October, 2017

Announcement

- Assignment 4 will be released tomorrow morning: Due on October 31st

Outline

- Review
- Introduction
- Delta method
- The χ^2 Distribution
- The t Distribution
- The Sample Mean and Sample Variance

Review

- Assessment of Estimators
  - Sampling distribution: distribution of $\hat{\theta}_n$
  - Consistency: $\hat{\theta} \to \theta$ in probability. Remark: MLE and MM are both consistent
  - Unbias: $E(\hat{\theta}_n) = \theta$, for any n
  - Variance: $\text{Var}(\hat{\theta}_n)$ is small
  - Mean Squared Error: $\text{MSE}(\hat{\theta}_n) = E(\hat{\theta}_n - \theta)^2 = \text{Bias}^2 + \text{Var}$
  - Remark: MSE converges to 0 $\Rightarrow \hat{\theta}_n$ is consistent
- Cramer-Rao Lower Bound
  - For any unbiased estimator $\hat{\theta}_n$,

$$\text{Var}(\hat{\theta}_n) \geq 1/(nI(\theta)), \qquad \text{any } n.$$

  - Efficiency: $(nI(\theta))^{-1}/\text{Var}(\hat{\theta}_n)$
  - If $\text{Var}(\hat{\theta})_n = 1/(nI(\theta))$ ($\hat{\theta}_n$ has efficiency 1), then $\hat{\theta}_n$ is *efficient.*

Introduction

Learning outcomes

- Questions to Address: What is delta method ⋆ Apply the distributions and delta method to data ⋆ The definition of χ^2 distribution ⋆ The definition of t Distribution ⋆ Relationship between normal, χ^2 and t distributions

Concept & Terminology

- Consistency ⋆ Slutsky's Theorem ⋆ Delta Method
- χ^2-distribution ⋆ χ^2-table
- t-distribution ⋆ t-table
- Sample mean ⋆ Sample variance

Mandatory Reading

Section 4.6, Section 6.1 - Section 6.3

Estimators for $N(\mu, \sigma^2)$

For normal r.v.'s $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ with unknown parameters (μ, σ^2), recall the MLE's are

$$\hat{\mu} = \bar{X}_n, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$E(\hat{\mu}) = E(\bar{X}_n) = \mu$, $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$; $\hat{\mu}$ is unbiased, but $\hat{\sigma}^2$ is not

- We adjust $\hat{\sigma}^2$ as

$$S_n^2 = \frac{n}{n-1}\cdot\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$$

  where $E(S_n^2) = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2$, unbiased!
- This is a generally used method to find unbiased estimators
- The two new estimators are generally used in practical, even without normal assumption. We call them as
  - Sample mean: $\bar{X}_n$;
  - Sample variance: $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

Sample mean and Sample variance with normal

For normal i.i.d. r.v.'s $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ with unknown parameters (μ, σ^2), there are

- Unbiased Estimators:
  Sample mean: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$
  Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

- Sampling distribution:
  Sample mean: $\bar{X}_n \sim N(\mu, \sigma^2/n)$
  Sample variance: approximate distribution

- Sample mean and sample variance are generally used even when $X_1, \cdots, X_n$ does not follow normal distribution
- Properties for Sample mean and Sample variance for non-normal data?

Consistency

1. According to WLLN, sample mean $\bar{X}_n$ is consistent
2. If $X_1, \cdots, X_n$ are i.i.d. r.v.'s, and we define

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

- If $X_1, \cdots, X_n$ are i.i.d. normal r.v.'s, S^2 is an unbiased and consistent estimator for σ^2
- If $X_1, X_2, \cdots$ are i.i.d. r.v.'s with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then a suffient condition that $S_n^2 \to \sigma^2$ is $\mathrm{Var}(S_n^2) \to 0$ as $n \to \infty$

Proof. $E(S_n^2) = E(\frac{1}{n-1}[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2]) = \frac{1}{n-1}(nEX_1^2 - nE\bar{X}^2)$

$$= \frac{1}{n-1}(n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)) = \sigma^2$$

So, $\mathrm{MSE}(S_n^2) = \mathrm{Var}(S_n^2)$, and $\mathrm{Var}(S_n^2) \to 0$ is equivalent with $\mathrm{MSE}(S_n^2) \to 0$, which is sufficient for consistency.

Consistency of S

If S_n^2 is consistent, then
$S_n = \sqrt{S_n^2} = h(S_n^2)$ is a consistent estimator of σ, such that

$$S_n \xrightarrow{P} \sigma$$

- Suppose that $X_1, X_2, \cdots$ converges in probability to a r.v. X and that h is a continuous function. Then $h(X_1)$, $h(X_2)$, $\cdots$ converges in probability to $h(X)$
- S_n is, in fact, a biased[1] estimator of σ, but the bias disappears asymptotically

[1] $E(S_n) \neq \sigma$

Normal approximation with estimated variance–CLT

For CLT, as the true variance σ is unknown, we usually use sample variance S_n^2 instead

$$\text{CLT: } \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to N(0,1)$$

Use S_n instead of the unknown σ

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sigma}{S_n}\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to N(0,1) \quad \textit{hopefully}$$

Does this convergence hold?

Slutsky's Theorem

If $X_n \xrightarrow{d} X$ in distribution and $Y_n \xrightarrow{P} a$, a is a constant, then

1. $Y_n X_n \xrightarrow{d} aX$
2. $Y_n + X_n \xrightarrow{d} X + a$

$$\frac{\sigma}{S_n} \xrightarrow{P} 1$$

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

$$\text{The product:} \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$$

Suppose $E(X) = \mu \neq 0$, and we want to estimate a function $g(\mu)$, where should we start?

- Taylor:

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu)$$

- A rough estimator of $g(\mu)$ can be

$$g(X)$$

  and we have

$$E(g(X)) \approx g(\mu) \text{ 2nd term disappears}$$

$$\text{Var}\left(g(X)\right) \approx [g'(\mu)]^2 \text{Var}(X) \text{ 1st term does not contribute}$$

- if $g(\mu) = 1/\mu$, then

$$E(\frac{1}{X}) \approx \frac{1}{\mu}, \quad \text{Var}(\frac{1}{X}) \approx (\frac{1}{\mu})^4 \text{Var}(X)$$

Estimating the odds ratio

Consider the population odds in Bernoulli(p)

$$\frac{p}{1-p}$$

Given data $X_1, X_2, \cdots, X_n$, we had $\hat{p} = \bar{X}$, and $\hat{p} \xrightarrow{P} p$ from WLLN. Try to replace p with $\hat{p}$ to estimate the *odds*

$$\frac{\hat{p}}{1-\hat{p}}$$

Question: what is the variance of $\frac{\hat{p}}{1-\hat{p}}$ (the new random variable)

Solution:

$$\mathrm{Var}(\frac{\hat{p}}{1-\hat{p}}) \approx [g'(p)]^2\mathrm{Var}(\hat{p}) = [(\frac{p}{1-p})']^2\mathrm{Var}(\hat{p})$$

$$= [\frac{1}{(1-p)^2}]^2\frac{p(1-p)}{n} = \frac{p}{n(1-p)^3}$$

As $\hat{p}$ is unbiased, meaning $E(\hat{p}) = p$, one might ask how good is the variance approximation?

Delta Method

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0, Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, [g'(\theta)]^2\sigma^2)$$

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \text{Remainder}$$

$$\sqrt{n}[g(Y_n) - g(\theta)] = g'(\theta)\sqrt{n}(Y_n - \theta)$$

$$g'(\theta)\sqrt{n}(Y_n - \theta) \xrightarrow{d} g'(\theta)N(0, \sigma^2) = N(0, [g'(\theta)]^2\sigma^2)$$

By Slutsky and note that Remainder $\xrightarrow{P} 0 \quad (Y_n \xrightarrow{P} \theta)$

Delta Method

The statement of the Delta Method allows for great generality of sequences Y_n satisfying the CLT, and also the generality of MLE. Typically, $g(\hat{\theta}_{MLE})$ is the MLE of $g(\theta)$, and the asymptotic distribution can be found through Delta Method and the asymptotic distribution of MLE (fisher information).

Continuation of example $g(\mu) = 1/\mu$

Suppose that we have a sample from X, with sample mean $\bar{X}$. Note that $E(\bar{X}) = \mu$, then for $\mu \neq 0$, we have

$$\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\mu}) \xrightarrow{d} N(0, (\frac{1}{\mu})^4 \mathrm{Var}(X_1))$$

- How to use it? replace everything by estimation.

$$(\frac{1}{\mu})^4 \widehat{\mathrm{Var}(X_1)} \approx (\frac{1}{\bar{X}})^4 S^2.$$

- The above is equivalent with

$$\frac{\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\mu})}{(\frac{1}{\bar{X}})^2 S} \xrightarrow{d} N(0,1)$$

- Proof. $\bar{X}$, S^2 are consistent estimator (WLLN), apply Slutsky's Theorem, to have that for $\mu \neq 0$,

$$\frac{\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\mu})}{(\frac{1}{\bar{X}})^2 S} = \frac{(\frac{1}{\mu})^2 \sigma}{(\frac{1}{\bar{X}})^2 S} \cdot \frac{\sqrt{n}(\frac{1}{\bar{X}} - \frac{1}{\mu})}{(\frac{1}{\mu})^2 \sigma} \xrightarrow{d} N(0,1)$$

Exact Distribution of Sample Variance

- The behavior of S_n^2 when $X_1, X_2, \cdots, X_n$ are i.i.d. r.v.'s is clear
- What if $X_1, X_2, \cdots, X_n$ are normal r.v.'s?
- Exact distribution of S_n^2?

Definition

- If $Z \sim N(0,1)$, the distribution of $U = Z^2$ is called the *chi-square distribution with 1 degree of freedom (df)*; we write $U \sim \chi^2_1$
- If $U_1, \cdots, U_n \sim \chi^2_1$ are independent, the distribution of $V = U_1 + \cdots + U_n$ is called the *chi-square distribution with n degree of freedom (df)*; we write $V \sim \chi^2_n$

- Obviously, if $X_1, X_2, \cdots, X_n \sim N(0,1)$ i.i.d, then $\sum_{i=1}^n X_i^2$ follows χ^2_n distribution
- The degree of freedom, equivalently as the number of independent normal r.v.'s in the summation, must be a positive integer

Density of χ^2_n and Properties

- The PDF of $V \sim \chi^2_n$ is

$$f_V(v) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} v^{n/2-1} e^{-v/2}, & v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- $E(V) = n$; $\text{Var}(V) = 2n$
- Sum of independent chi-square r.v.'s:

$$\chi^2_m + \chi^2_n = \chi^2_{m+n}$$

- Relationship with other distributions: $\chi^2_n = \text{Gamma}(n/2, 1/2)$

Let $X \sim N(0,1)$ and $Y = X^2$. Let $f(y)$ and $F(y)$ denote, respectively, the PDF and the CDF of Y; let $\phi(x)$ and $\Phi(x)$ denote the PDF and the CDF of X. Then for $y > 0$

$$\begin{aligned} F(y) = P(Y \le y) = P(X^2 \le y) &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \end{aligned}$$

We know $f(y) = F'(y)$ and $\phi(x) = \Phi'(x)$, it follows from the chain rule for derivatives that

$$\begin{aligned} f(y) = F'(y) &= (\Phi(y^{1/2}) - \Phi(-y^{1/2}))' \\ &= \phi(y^{1/2}) \left(\frac{1}{2} y^{-1/2}\right) + \phi(-y^{1/2}) \left(\frac{1}{2} y^{-1/2}\right) \\ &= (2\pi)^{-1/2} e^{-y/2} \left(2 \cdot \frac{1}{2} y^{-1/2}\right) \ (\phi(x) \text{ is symmetric}) \\ &= \frac{1}{(2\pi)^{1/2}} y^{-1/2} e^{-y/2} \end{aligned}$$

mgf of χ_1^2 is

$$M(t) = (1-2t)^{-1/2}, \ \ t < 1/2$$

Proof:

$$\begin{aligned}
E(e^{tv}) &= \int_{-\infty}^{\infty} e^{tv} f(v) dv = \frac{1}{2^{1/2}\Gamma(1/2)} \int_0^{\infty} e^{tv} v^{1/2-1} e^{-v/2} dv \\
&= c \int_0^{\infty} v^{1/2-1} e^{-(1/2-t)v} dv \ (c = \frac{1}{2^{1/2}\Gamma(1/2)}) \\
&= c \int_0^{\infty} \left(\frac{2}{1-2t} y\right)^{1/2-1} e^{-y} (\frac{2}{1-2t}) dy \ (y = (1/2-t)v) \\
&= c \int_0^{\infty} \left(\frac{2}{1-2t}\right)^{1/2} y^{-1/2} e^{-y} dy \\
&= c \left(\frac{2}{1-2t}\right)^{1/2} \int_0^{\infty} y^{-1/2} e^{-y} dy = c \left(\frac{2}{1-2t}\right)^{1/2} \Gamma(1/2) \\
&= (1-2t)^{-1/2}
\end{aligned}$$

Therefore, mgf of χ_n^2 is

$$M(t) = \prod_{i=1}^{n} M_i(t) = (1-2t)^{-n/2}, \ \ t < 1/2$$

From mgf

$$M(t) = (1-2t)^{-n/2}, \;\; t < 1/2$$

$$M^{'}(t) = (-n/2)(1-2t)^{-n/2-1} \cdot (-2) = n(1-2t)^{-n/2-1}$$

$$M^{''}(t) = \left(n(1-2t)^{-n/2-1}\right)^{'} = n(n+2)(1-2t)^{-n/2-2}$$

We calculate the first $(E(X))$ and second $(E(X^2))$ moments

$$E(X) = M^{'}(0) = n$$

$$E(X^2) = M^{''}(0) = n(n+2)$$

Therefore

$$\mathrm{Var}(X) = E(X^2) - E^2(X) = 2n$$

Density of χ_n^2 and Properties - cont'd

From mgf, if $X \sim \chi_n^2$, $Y \sim \chi_m^2$ and $X \perp Y$

$$M_X(t) = (1-2t)^{-n/2}, \;\; t < 1/2; \qquad M_Y(t) = (1-2t)^{-m/2}, \;\; t < 1/2$$

Therefore,

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= (1-2t)^{-n/2} \cdot (1-2t)^{-m/2} \\ &= (1-2t)^{-(n+m)/2} \end{aligned}$$

As a conclusion, $X + Y \sim \chi_{n+m}^2$

χ^2 and Normal

- If $X_1, \cdots, X_n \sim N(0,1)$ i.i.d, $\sum_{i=1}^n X_i^2 \sim \chi_n^2$
- Recall that if $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$
- Therefore, if $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ i.i.d, $\sum_{i=1}^n (\frac{X_i-\mu}{\sigma})^2 \sim \chi_n^2$
- If $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ i.i.d., yet μ is unknown and estimated by the sample mean $\bar{X}_n$, then $\sum_{i=1}^n (\frac{X_i-\bar{X}_n}{\sigma})^2 \sim \chi_{n-1}^2$, still a χ^2 distribution, but the degrees of freecom change from n to $n-1$ $\Leftrightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

$$W = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma} \right)^2$$

expand it

$$\underbrace{W}_{\chi_n^2} = \sum_{i=1}^{n} \left(\frac{(X_i - \bar{X})}{\sigma} \right)^2 + \sum_{i=1}^{n} \left(\frac{(\bar{X} - \mu)}{\sigma} \right)^2 + \overbrace{2 \left(\frac{\bar{X} - \mu}{\sigma^2} \right) \sum_{i=1}^{n} (X_i - \bar{X})}^{0}$$

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

It can be proved that s^2 is independent with $\bar{X}$. With mgf, and also $W \sim \chi_n^2$, so $\frac{(n-1)s^2}{\sigma^2}$ is χ_{n-1}^2.

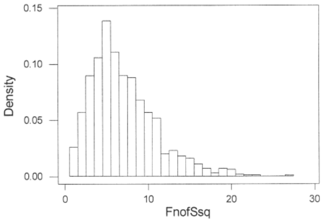
Example: χ^2 Distribution $(n = 8)$

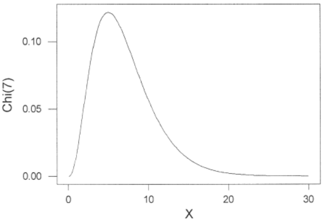
- Would like to see the sampling distribution of χ^2 $(n = 8)$
- Generate 1000 samples with each of size 8 from $N(100, 256)$, and calculate

$$\chi^2 = \frac{(8-1)s^2}{\sigma^2} = \frac{7s^2}{\sigma^2} = \frac{\sum_{i=1}^{8}(X_i - \bar{X})^2}{256}$$

Row	X1	X2	X3	X4	X5	X6	X7	X8	FnofSsq
1	98	77	96	116	122	89	100	91	5.7651
2	104	104	107	106	96	106	74	92	3.4917
3	81	108	100	108	122	107	97	110	3.8862
4	94	95	93	121	93	113	114	94	3.7690
5	111	91	104	111	84	98	100	101	2.3438
6	106	91	87	94	98	61	107	107	6.4253
7	101	117	121	111	80	125	129	84	9.3594
8	102	103	95	91	82	112	83	85	3.2222
9	71	94	107	104	139	112	103	95	10.0112
10	100	87	119	89	93	68	124	108	9.1641
11	109	124	109	101	91	81	104	125	6.1719
12	101	87	113	72	83	85	114	102	6.2729
13	69	96	105	87	108	113	128	125	10.5190
14	108	102	102	114	103	100	77	95	3.2456
... and so on ...									
995	129	137	112	126	95	91	98	97	8.7729
996	85	118	84	109	97	96	98	110	3.9604
997	107	82	123	96	115	137	100	109	7.8042
998	115	82	92	86	137	90	97	68	12.4019
999	107	115	122	96	110	80	142	111	8.9761
1000	100	130	102	59	90	82	115	97	12.3628

- The histogram looks similar to that of the density curve of a chi-square random variable with 7 degrees of freedom





Example: Acid Concentration in Cheese

The variation in concentrations of chemicals like lactic acid can lead to variation in the taste of cheese. Suppose that we model the concentration of lactic acid in several chunks of cheese as independent normal random variables with mean μ and variance σ^2. We are interested in how much these concentrations differ from the value μ.

Solution: Let $X_1, \cdots, X_k$ be the concentrations in k chunks, and let $Z_i = (X_i - \mu)/\sigma$. Then

$$Y = \frac{1}{k}\sum_{i=1}^{k}(X_i - \mu)^2 = \frac{\sigma^2}{k}\sum_{i=1}^{k} Z_i^2$$

is one measure of how much the k concentrations differ from μ. We might need to compute $P(Y \leq u^2)$. The distribution of

$$W = \frac{kY}{\sigma^2} \sim \chi_k^2$$

Hence,

$$P(Y \leq u^2) = P(W \leq ku^2/\sigma^2)$$

Example: Acid Concentration in Cheese

Recall,

$$P(Y \leq u^2) = P(W \leq ku^2/\sigma^2)$$

Suppose that $\sigma^2 = 0.09$, and we are interested in $k = 10$ cheese chunks. Furthermore, suppose that $u = 0.3$ is the critical difference of interest. We have

$$P(Y \leq 0.3^2) = P\left(W \leq \frac{10 \times 0.09}{0.09}\right) = P(\overbrace{W}^{\chi^2_{10}} \leq 10) \approx 0.56$$

So there is a 44 percent chance that the average squared difference between lactic acid concentration and mean concentration in 10 chunks will be more than the desired amount. If this probability is too large, the manufacturer might wish to invest some effort in reducing the variance of lactic acid concentration

- With χ^2 distribution, if $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$ i.i.d, we know the *exact distribution* for S_n^2
- What's more, according to Slutsky's Theorem, we have

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

- What is the *exact distribution* for $\sqrt{n}\frac{\bar{X}-\mu}{S_n}$?

The Real *Student* William Sealy Gosset

- Gosset had almost all his papers including *The probable error of a mean* published (1907) in Pearson's journal *Biometrika* under the pseudonym *Student*.
- It was, however, not Pearson but Ronald A. Fisher who appreciated the importance of Gosset's small-sample work.

William Sealy Gosset

Student in 1908

(2) "STUDENT" AS STATISTICIAN

By E. S. PEARSON

For many years after the publication of his first paper in *Biometrika*, in 1907, the name of "Student" was associated in statistical circles with an atmosphere of romance. Those who knew him only through his written contributions must often have wondered who was this unassuming man, content to remain anonymous, who wrote so clearly and simply on so wide a range of fundamental topics. To those of us who came into touch with him personally, the knowledge that "Student" was W. S. Gosset did not altogether dispel that romantic impression. Here, in London, he would pay us visits from time to time at the old Biometric Laboratory on his way to Euston station to catch the Irish mail;

We define a parametric distribution called the t distribution by dividing a standard normal r.v. by the square root of an independently chi-square r.v. with n df scaled by $1/n$

Definition

If $Z \sim N(0,1)$ and $U \sim \chi^2_n$ are independent, the distribution of $T = Z/\sqrt{U/n}$ is called t distribution with n degree of freedom, where n is a positive integer; we write $T \sim t_n$

pdf of a t_n r.v.

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

- $E(T) = 0, n > 1$ and $\mathrm{Var}(T) = \frac{n}{n-2}, n > 2$

- t has no mgf because it does not have moments of all orders.
- If there are n degrees of freedom, then there are only $n-1$ moments.
- t_1 has no mean; t_2 has no variance

Find the dist. of $T = U/\sqrt{V/n}$, $U \sim N(0,1)$ and $V \sim \chi^2_n$ are independent:
The joint pdf of U and V is

$$f_{U,V}(u,v) = \frac{1}{(2\pi)^{1/2}} e^{-u^2/2} \cdot \frac{1}{\Gamma(n/2)2^{n/2}} v^{(n/2)-1} e^{-v/2}, -\infty < u, v < \infty$$

Make the transformation

$$t = \frac{u}{\sqrt{v/n}}, \quad w = v$$

The Jacobian is $(w/n)^{1/2}$, and the marginal pdf of T is

$$f_T(t) = \int_0^\infty f_{U,V}(t(\frac{w}{n})^{1/2}, w)(w/n)^{1/2} dw$$

$$f_T(t) = \frac{1}{(2\pi)^{1/2}} \frac{1}{\Gamma(n/2)2^{(n/2)}} \int_0^\infty e^{(-1/2)t^2 w/n} w^{(n/2)-1} e^{-w/2} (w/n)^{1/2} dw$$

$$\frac{1}{(2\pi)^{1/2}} \frac{1}{\Gamma(n/2)2^{(n/2)} n^{1/2}} \underbrace{\int_0^\infty e^{-(1/2)(1+t^2/n)w} w^{((n+1)/2)-1} dw}_{\text{integrand of } gamma((n+1)/2, 2/(1+t^2/n))}$$

$$\int_0^\infty e^{-(1/2)(1+t^2/n)w} w^{((n+1)/2)-1} dw = \Gamma((n+1)/2)[\frac{2}{1+t^2/n}]^{(n+1)/2}$$

recall gamma pdf:

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

A Computer Simulation of t Distribution

Generate 1000 t_3 Samples

Following the definition

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

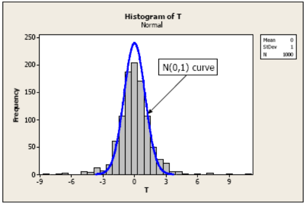
let's randomly generate 1000 standard normal values (Z) and 1000 chi-square(3) values (U). Then, the above definition tells us that, if we take those randomly generated values, calculate:

$$T = \frac{Z}{\sqrt{U/3}}$$

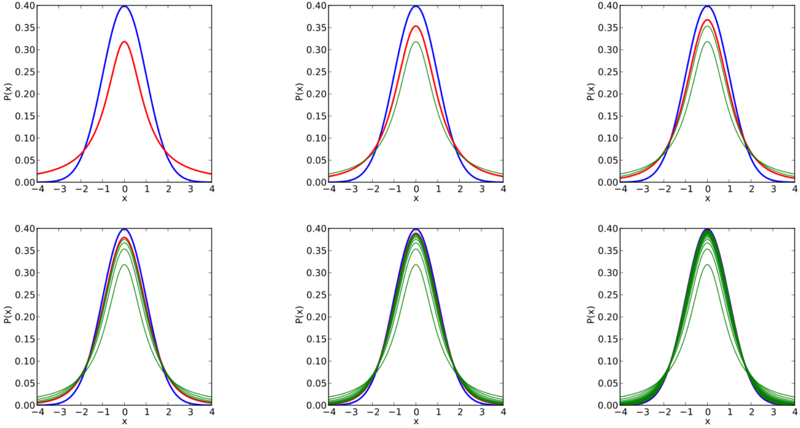
and create a histogram of the 1000 resulting T values, we should get a histogram that looks like a t distribution with 3 degrees of freedom.

Row	Z	Chisq(3)	T(3)
1	-2.60481	10.2497	-1.4092
2	2.92321	1.6517	3.9396
3	-0.48633	0.1757	-2.0099
4	-0.48212	3.8283	-0.4268
5	-0.04150	0.2422	-0.1461
6	-0.84225	0.0903	-4.8544
7	-0.31205	1.6326	-0.4230
8	1.33068	5.2224	1.0086
9	-0.64104	0.9401	-1.1451
10	-0.05110	2.2632	-0.0588
11	1.61601	4.6566	1.2971
12	0.81522	2.1738	0.9577
13	0.38501	1.8404	0.4916
14	-1.63426	1.1265	-2.6669
... and so on ...			
994	-0.18942	3.5202	-0.1749
995	0.43078	3.3585	0.4071
996	-0.14068	0.6236	-0.3085
997	-1.76357	2.6188	-1.8876
998	-1.02310	3.2470	-0.9834
999	-0.93777	1.4991	-1.3266
1000	-0.37665	2.1231	-0.4477

$$T(3) = \frac{-2.60481}{\sqrt{10.2497/3}} = -1.4092$$



Density of the t-distribution (red) for 1, 2, 3, 5, 10, and 30 degrees of freedom compared to the standard normal distribution (blue) [2]



[2]figures are obtained from wiki

As the degrees of freedom n increases, the t density curve gets closer and closer to the standard normal curve.

Properties of t Distribution

- The support is $\infty < t < \infty$.
- The probability distribution is symmetric about $t = 0$.
- The probability distribution appears is bell-shaped.
- The density curve looks like a standard normal curve, but the tails of the t-distribution are "heavier" than the tails of the normal distribution. That is, we are more likely to get extreme t-values than extreme z-values.
- As the degrees of freedom n increases, the t-distribution approaches the standard normal z-distribution.

Computing the probability for χ^2, t r.v.s

- Probabilities of X, which has a distribution as χ^2, t, are defined by

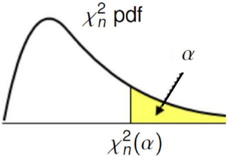
$$P(c \leq X \leq d) = \int_c^d f(x)dx$$

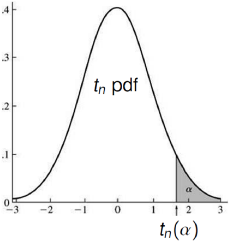
- Integrals of all two different $f(x)$ (χ^2_n or t_n) as integrands cannot be explicitly calculated; there are no closed-form solutions.
- Use χ^2-table (A8) and t-table (A9) for computing probabilities

Notation for percentiles of the χ^2, t, Z distributions ($0 \leq \alpha \leq 1$)

Definition

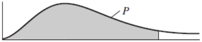
Let $\chi^2_n(\alpha)$ (resp. $t_n(\alpha)$ or $Z(\alpha)$) denotes the point beyond which the χ^2_n (resp. t_n or Z) r.v., has probability α. Equivalently, it is also $100(1-\alpha)$-th percentile of the χ^2_n (resp. t_n or Z) distribution





How to Use the χ^2 Table

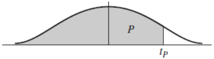
TABLE 3 Percentiles of the χ^2 Distribution—Values of χ^2_P Corresponding to P



df	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.000039	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

How to Use the t Table

TABLE 4 Percentiles of the t Distribution



df	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	.325	.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.169
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787

Steps you should use in using the χ^2 table (similarly for t table)

1. Find the row that corresponds to the relevant degrees of freedom, df.
2. Find the column headed by the probability of interest... whether it's 0.005, 0.01, 0.025, 0.05, 0.10, 0.90, 0.95, 0.975, 0.99, or 0.995. Determine the chi-square value where the df row and the probability column intersect.

χ^2

1. From Table 3 (page A8) of textbook, we have the following χ^2 percentiles

$$P(\chi_1^2 < .000039) = .005 \Leftrightarrow \chi_1^2(0.995) = 0.000039^3$$
$$P(\chi_5^2 < 1.61) = .10 \Leftrightarrow \chi_5^2(.90) = 1.61$$
$$P(\chi_5^2 < 12.83) = .975 \Leftrightarrow \chi_5^2(.025) = 12.83$$
$$P(\chi_5^2 < 16.75) = .995 \Leftrightarrow \chi_5^2(.005) = 16.75$$

[3] the table gives the cumulative probability and the definition of $\chi_n^2(\alpha)$ is reverse!

t

1. From Table 4 (page A9) of textbook, we have the following t percentiles

$$
\begin{aligned}
P(t_1 < .325) &= .60 \Leftrightarrow t_1(.40) = .325^4 \\
P(t_9 < 2.262) &= .975 \Leftrightarrow t_9(.025) = 2.262 \\
P(t_9 < -2.262) &= P(t_9 > 2.262) = 1 - P(t_9 < 2.262) \\
&= 1 - .975 = .025 \Leftrightarrow t_9(.975) = -2.262 \\
P(-.261 < t_9 < 2.262) &= P(t_9 < 2.262) - P(t_9 < -.261) \\
&= .975 - (1 - .60) = .575
\end{aligned}
$$

[4] the table gives the cumulative probability and the definition of $t_n(\alpha)$ is reverse!

Assume $X_1, \cdots, X_n$ be an independent sample from $N(\mu, \sigma^2)$, and

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

Distribution of $\bar{X}$ of iid Normal Samples

The distribution of $\bar{X}_n$ is $N(\mu, \frac{\sigma^2}{n})$

Independence between $\bar{X}$ and S_n^2 of iid Normal Samples

$\bar{X}$ and S_n^2 are *independent*

Distribution of s^2 of iid Normal Samples

The distribution of $(n-1)S_n^2/\sigma^2$ is χ^2_{n-1}

Summary of $\bar{X}_n$ and S_n^2– cont'd

A t distribution built on $\bar{X}_n$ and S_n^2

- Let $\bar{X}_n, S_n^2$ be the sample mean and sample variance defined from iid samples from $N(\mu, \sigma^2)$. Then

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

- Follows by definition of a t distribution as the ratio

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}} = \frac{\overbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}^{N(0,1)}}{\underbrace{\sqrt{\frac{(n-1)s^2/\sigma^2}{(n-1)}}}_{\chi^2_{n-1}/(n-1)}}$$

Example 1 (textbook page 280)

We found the mle's of μ and σ^2 from an iid normal sample:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Build the $100(1-\alpha)\%$ CI for μ and σ.

A confidence interval for μ is based on

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

where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Let $t_{n-1}(\alpha/2)$ denote the point beyond which the t distribution with $n-1$ df has probability $\alpha/2$. Then, by definition,

$$P\left(-t_{n-1}(\alpha/2) \le \frac{\sqrt{n}(\bar{X} - \mu)}{S} \le t_{n-1}(\alpha/2)\right) = 1 - \alpha$$

After some manipulation,

$$P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2) \le \mu \le \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2)\right) = 1 - \alpha$$

The confidence interval for μ is

$$(\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2), \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha/2))$$

Similarly, we turn to a confidence interval for σ^2

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

where $\chi^2_m(\alpha)$ denote the point beyond which the chi-square distribution with m df has probability α. It follows from a definition (chi-square is not symmetric)

$$P\left(\chi^2_{n-1}(1 - \alpha/2) \le \frac{n\hat{\sigma}^2}{\sigma^2} \le \chi^2_{n-1}(\alpha/2)\right) = 1 - \alpha$$

Example: χ^2-distribution for CI

After some manipulation,

$$P\left(\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)} \leq \sigma^2 \leq \frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}\right) = 1-\alpha$$

Therefore, the $100(1-\alpha)\%$ confidence interval is

$$\left(\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)}, \frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$$

Remark: this interval is not symmetric about $\hat{\sigma}^2$, meaning it is not of the form $\hat{\sigma}^2 \pm c$