50.034 - Introduction to Probability and Statistics

Week 10 – Lecture 18

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 18 Assigned Readings: 8.4, 8.5, 8.7

Outline of Lecture

- t-distribution
- t-distribution versus normal distribution
- Confidence interval
- Unbiased estimation

Example 1

Let $\{X_1, \ldots, X_{100}\}$ be a random sample consisting of 100 **normal** observable R.V.'s with **unknown mean** μ and known variance 25.

Question: Can we find a statistic T of $\{X_1, \ldots, X_{100}\}$ such that $\Pr(T \leq \mu) = 0.9$?

- **Recall:** A statistic is a function of observable R.V.'s.
- In other words, can we find a R.V. $T = f(X_1, \ldots, X_{100})$ such that the event $\{T \leq \mu\}$ has a 90% probability of occurring, even though μ is unknown to us?
- The challenge is to find T such that no matter what the value of μ is, there will always be a probability of 90% that the event $\{T \leq \mu\}$ occurs.

Example 1 - Solution

Since $X_1, \ldots, X_{100}$ are normal, here are some facts we know:

- The sample mean $\overline{X}_{100}$ is normal with mean μ and variance $\frac{25}{100} = \frac{1}{4}$. (So the standard deviation of $\overline{X}_{100}$ is $\frac{1}{2}$.)
- Hence, $\frac{\overline{X}_{100}-\mu}{\frac{1}{2}} = 2(\overline{X}_{100} - \mu)$ is a standard normal R.V.

Let $\Phi(z)$ be the standard normal cdf. From the standard normal distribution table, the closest value of z satisfying $\Phi(z) = 0.9$ is $z = 1.28$.

Thus,

$$\begin{aligned} 0.9 &= \Pr(2(\overline{X}_{100} - \mu) \leq 1.28) = \Pr(\overline{X}_{100} - \mu \leq 0.64) \\ &= \Pr(\overline{X}_{100} - 0.64 \leq \mu). \end{aligned}$$

Therefore, we can let T be the statistic $\overline{X}_{100} - 0.64$.

Example 2

(Extension of Example 1 to the case of two statistics)

Let $\{X_1, \ldots, X_{100}\}$ be a random sample consisting of 100 **normal** observable R.V.'s with **unknown mean** μ and known variance 25.

Question: Can we find two statistics T_1, T_2 of $\{X_1, \ldots, X_{100}\}$ such that $\Pr(T_1 \leq \mu \leq T_2) = 0.9$?

- In other words, can we find two R.V.'s $T_1 = f_1(X_1, \ldots, X_{100})$ and $T_2 = f_2(X_1, \ldots, X_{100})$ such that the event

$$\{T_1 \leq \mu\} \cap \{T_2 \geq \mu\}$$

has a 90% probability of occurring, even though μ is unknown?

- The challenge is to find T_1 and T_2 such that no matter what the value of μ is, there will always be a probability of 90% that the event $\{T_1 \leq \mu\} \cap \{T_2 \geq \mu\}$ occurs.

Example 2 - Solution

As observed in Example 1, $2(\overline{X}_{100} - \mu)$ is a standard normal R.V.

Let $\Phi(z)$ be the standard normal cdf. Given any real number $r > 0$,

$$\begin{aligned}
&\Pr(\overline{X}_{100} - r \leq \mu \leq \overline{X}_{100} + r) = \Pr(-r \leq \overline{X}_{100} - \mu \leq r) \\
&= \Pr(-2r \leq 2(\overline{X}_{100} - \mu) \leq 2r) = \Phi(2r) - \Phi(-2r) \\
&= \Phi(2r) - (1 - \Phi(2r)) = 2 \cdot \Phi(2r) - 1.
\end{aligned}$$

So if we equate $\Pr(\overline{X}_{100} - r \leq \mu \leq \overline{X}_{100} + r) = 0.9$, then it means that $\Phi(2r) = 0.95$.

From the standard normal distribution table, the closest value of z satisfying $\Phi(z) = 0.95$ is $z = 1.64$. Hence $2r = 1.64$, which implies $r = 0.82$.

Therefore, we can let T_1 and T_2 be the statistics $\overline{X}_{100} - 0.82$ and $\overline{X}_{100} + 0.82$ respectively.

Dealing with unknown variance

Let $\{X_1, \ldots, X_n\}$ be a random sample of **normal** observable R.V.'s with **unknown mean** μ.

Question: Given some $0 < p < 1$, can we find two statistics T_1, T_2 of $\{X_1, \ldots, X_n\}$ such that $\Pr(T_1 \leq \mu \leq T_2) = p$?

Answer: If the variance of each X_i is known, then we can easily adapt the method from Example 2 to find T_1 and T_2.

$$\boxed{\text{What if we do not know the variance of } X_i\text{?}}$$

- Perhaps we could approximate σ by the sample variance $\hat{\sigma}^2$?
  - But $\sigma^2 \approx \hat{\sigma}^2$ only if the sample size n is large.
  - For large n, the law of large numbers already gives $\overline{X}_n \approx \mu$.
  - What if the sample size is small?
- Given a fixed p, can we find some statistics T_1 and T_2 such that no matter what the values of μ and σ^2 are, the event $\{T_1 \leq \mu\} \cap \{T_2 \geq \mu\}$ will always occur with probability p?
- Can we find such T_1 and T_2 even in the case when the sample size n is small?

Dealing with unknown variance (continued)

Let $\{X_1, \ldots, X_n\}$ be a random sample of **normal** observable R.V.'s with **unknown mean** μ and **unknown variance** σ^2.

Important Breakthrough in Statistics:
Given any fixed p $(0 < p < 1)$ and any sample size n, it is possible to find statistics T_1 and T_2 that satisfy $\Pr(T_1 \leq \mu \leq T_2) = p$, no matter what the values of μ and σ^2 are.

- **Note:** We need the assumption that $X_1, \ldots, X_n$ are normal!
  - We cannot possibly find T_1 and T_2 if we know absolutely nothing about the random sample other than the value of n.
- T_1 and T_2 will not depend on the values of μ and σ^2.
- **Key Idea:** A new class of distributions called **t-distributions**.

Interpretation: T_1 can be interpreted as an "under-estimate" of μ, and T_2 can be interpreted as an "over-estimate" of μ. Together, the range from T_1 to T_2 would contain μ with probability p.

- **Huge practical implications:** Many real-world R.V.'s can be modeled as normal, both mean and variance are frequently unknown, and random samples are usually small!

Introduction to t-distribution

The t distributions were introduced by W. S. Gosset in 1908.

- "t-distribution" is also called "Student's t-distribution"

Interesting Back Story:

- Gosset published his research on the t-distribution under the pen name "Student".

- In 1908, he was the Head Experimental Brewer of Guinness.
  - His role was to test the quality and taste of Guinness beer, as well as invent new beer.

Image source: Jamt9000

- For fear of leaking trade secrets (e.g. beer recipe), Guinness did not allow any employee to publish research that contains the words "beer", "Guinness", or the employee's surname.
- Gosset actually used t-distributions for statistical inference on beer samples.

t-distribution

Let X be a continuous R.V.

Definition 1: We say X has a t-distribution if its pdf is given by:

$$f(x) = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi} \cdot \Gamma(\frac{m}{2})} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2} \quad \text{(for all } x\text{)}$$

for some positive integer m.

Definition 2: We say X has a t-distribution if

$$X = \frac{Z}{\sqrt{\frac{Y}{m}}},$$

for some positive integer m, where $Z \sim N(0,1)$, and $Y \sim \chi^2(m)$.

- Both definitions are equivalent. (See course textbook for a proof.)
- The positive integer m is called the degree of freedom.
- We say X has the t-distribution with m degrees of freedom.
    - We rarely say X is a t R.V., since this could be ambiguous.

Main Use: To model a normal R.V. with unknown mean and unknown variance after some transformation.

Unbiased sample variance

Let $\{X_1, \ldots, X_n\}$ be a random sample of observable R.V.'s.
Definition: The unbiased sample variance of $\{X_1, \ldots, X_n\}$ is

$$s_n^2(X_1, \ldots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

- Common notation for unbiased sample variance: s_n^2 or S_n^2.
- s_n^2 is also called the Bessel-corrected sample variance.
- $s_n = \sqrt{s_n^2}$ is called unbiased sample standard deviation.

In comparison, the biased sample variance of $\{X_1, \ldots, X_n\}$ (which we saw in Lecture 16) is

$$\hat{\sigma}^2(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is called biased sample standard deviation.
- Both $\hat{\sigma}^2$ and s_n^2 are commonly used, so be careful to state whether you mean biased or unbiased sample variance!

Note: $s_n^2 = \frac{n}{n-1} \hat{\sigma}^2$. (They are different by a factor of $\frac{n}{n-1}$.)

Main Theorem on t-distributions

Let $\{X_1, \ldots, X_n\}$ be a random sample of observable **normal** R.V.'s with mean μ and variance σ^2. Let $\overline{X}_n$ and s_n^2 be the sample mean and the **unbiased sample variance** respectively.

Important Theorem:
$\frac{\sqrt{n}(\overline{X}_n - \mu)}{s_n}$ has the t-distribution with $(n-1)$ degrees of freedom.

- In comparison, $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ has the standard normal distribution.

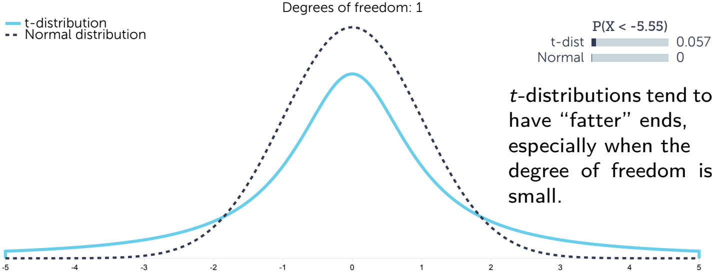
Interpretation:

- If σ^2 is known, then T_1, T_2 satisfying $\Pr(T_1 \leq \mu \leq T_2) = p$ can be found using the standard normal distribution.
- If σ^2 is unknown, then T_1, T_2 satisfying $\Pr(T_1 \leq \mu \leq T_2) = p$ could still be found using the t-distribution.

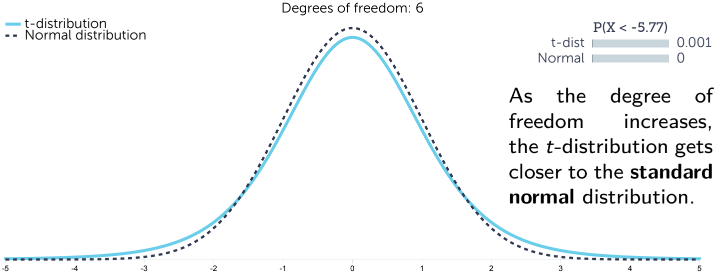
Note: (T_1, T_2) and (T_1, T_2) are different pairs of statistics!

- T_1, T_2 are computed with **less information** (σ^2 is unknown).
- However, if $n \to \infty$, then $s_n^2 \xrightarrow{\text{p}} \sigma^2$, hence (T_1, T_2) becomes closer to (T_1, T_2).

Shape of t-distribution



t-distributions tend to have "fatter" ends, especially when the degree of freedom is small.



As the degree of freedom increases, the t-distribution gets closer to the **standard normal** distribution.

Source: Kristoffer Magnusson (https://rpsychologist.com/d3/tdist/)

Useful properties of t-distribution

Suppose Z has the t-distribution with m degrees of freedom.
Let $f(z)$ be the pdf of Z, and let $F(z)$ be the cdf of Z.

Fact: The graph of $f(z)$ is symmetric about the point $z = 0$.

Fact: $F(-z) = 1 - F(z)$ for all real numbers z.

- This follows from the symmetry of the pdf $\phi(z)$.
- In other words, $\Pr(Z \le z) = \Pr(Z \ge -z)$ for all $z \in \mathbb{R}$.
- **Note:** $0 < F(z) < 1$ for all real numbers z.

Fact: $F^{-1}(p) = -F^{-1}(1 - p)$ for all real numbers $0 < p < 1$.

- To see why, let $z = F^{-1}(p)$ in previous fact, and apply F^{-1}.

Remarks on t-distribution

Special Case: t-distribution with one degree of freedom

- This special case is also called Cauchy distribution.
- If X has the Cauchy distribution, then $\mathbf{E}[X]$ does not exist!
  - (Lecture 5) For $\mathbf{E}[X]$ to exist, the following technical condition must hold:

    $$\int_0^\infty x \cdot f(x)\,dx < \infty \quad \text{or} \quad \int_{-\infty}^0 (-x) \cdot f(x)\,dx < \infty \quad \text{(or both).}$$

    But for Cauchy distribution, both integrals are infinite.

In general: If X has the t-distribution with m degrees of freedom, then the moment generating function of X does not exist!

- The k-th moment of X exists if $k < m$, but does not exist if $k \geq m$.
- For $m > 1$, the mean $\mathbf{E}[X]$ exists and equals 0.
- For $m > 2$, the variance $\operatorname{var}(X)$ exists and equals $\frac{m}{m-2}$.

t-distribution versus standard normal distribution

Theorem: For each $n \geq 1$, let X_n be the R.V. that has the t-distribution with n degrees of freedom. Then the infinite sequence of R.V.'s $X_1, X_2, X_3, \ldots$ converges in distribution to the standard normal R.V. Z, i.e. $X_n \xrightarrow{\text{d}} Z$.

- In other words, the asymptotic distribution of $X_1, X_2, X_3, \ldots$ is standard normal.

Recall: If $\{X_1, \ldots, X_n\}$ is a random sample of observable normal R.V.'s with mean μ and variance σ^2, and if $\overline{X}_n$ and s_n^2 are the sample mean and the unbiased sample variance respectively, then:

- $\frac{\sqrt{n}(\overline{X}_n - \mu)}{s_n}$ has t-distribution with $(n-1)$ degrees of freedom.
- $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ has the standard normal distribution.

Intuition: As $n \to \infty$, the unbiased sample variance s_n^2 approaches the "true" variance σ^2, so $\frac{\sqrt{n}(\overline{X}_n - \mu)}{s_n}$ would become approximately $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$. Therefore, for a sufficiently large degree of freedom, the t-distribution is approximately standard normal.

Computing the cdf of the t-distribution

If X has the t-distribution with m degrees of freedom, then its cdf is

$$F_X(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi} \cdot \Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{u^2}{m}\right)^{-(m+1)/2} du$$

- **Problem:** This integral has no closed form formula!
- Hence, the cdf $F_X(x)$ can only be computed approximately, using integral approximation methods (e.g. trapezoidal rule).

How your calculator/software package “computes” $\mathbf{F_X(x)}$ today:

- Approximations to $F_X(x)$ for various possible x and different degrees of freedom m are calculated beforehand, then stored in the calculator/software package in lookup tables.
- For any query, the stored value for the closest value found would be retrieved.
  - Similar to queries for normal distribution and χ^2 distribution.

Table for values of the t-distribution

(Before there were computers:) Computed values are stored as tables of values. For example, there is a table at the back of the course textbook called the "Table of the χ^2 Distribution".

- For example, if X has the t-distrubution with 9 degrees of freedom, then the closest value found for x satisfying $F_X(x) = 0.7$ is $x \approx 0.543$.

Table of the *t* Distribution

If X has a t distribution with m degrees of freedom, the table gives the value of x such that $\Pr(X \leq x) = p$.

m	$p = .55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	$\boxed{.543}$	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169

Example 3

Let $\{X_1, X_2, X_3, X_4\}$ be a random sample of observable normal R.V.'s with unknown mean μ and unknown variance σ^2. Let $\overline{X}_4$ and s_4^2 denote the sample mean and the unbiased sample variance respectively of $\{X_1, X_2, X_3, X_4\}$.

Question: Find two statistics T_1, T_2 of $\{X_1, X_2, X_3, X_4\}$, in terms of $\overline{X}_4$ and s_4, such that $\Pr(T_1 < \mu < T_2) = 0.8$ is satisfied.

Example 3 - Solution

Notice that $Z = \frac{\sqrt{4}(\overline{X}_4 - \mu)}{s_4} = \frac{2(\overline{X}_4 - \mu)}{s_4}$ has the t-distribution with 3 degrees of freedom. Let $F(z)$ be the cdf of Z. Then:

$$\begin{aligned}
\Pr(\overline{X}_4 - r < \mu < \overline{X}_4 + r) &= \Pr\left(-\frac{2r}{s_4} < \frac{2(\overline{X}_4 - \mu)}{s_4} < \frac{2r}{s_4}\right) \\
&= F\left(\frac{2r}{s_4}\right) - F\left(-\frac{2r}{s_4}\right) \\
&= F\left(\frac{2r}{s_4}\right) - \left(1 - F\left(\frac{2r}{s_4}\right)\right) \\
&= 2 \cdot F\left(\frac{2r}{s_4}\right) - 1.
\end{aligned}$$

Thus, $\Pr(\overline{X}_4 - r < \mu < \overline{X}_4 + r) = 0.8 \Leftrightarrow F\left(\frac{2r}{s_4}\right) = 0.9$.
From the table, $\frac{2r}{s_4} \approx 1.638$, hence $r \approx 0.819 s_4$, therefore we can let T_1 and T_2 be $\overline{X}_4 - 0.819 s_4$ and $\overline{X}_4 + 0.819 s_4$ respectively.

m	$p = .55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965
3	.137	.277	.424	.584	.765	.978	1.250	$\boxed{1.638}$	2.353	3.182	4.541
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365

Confidence Intervals of Parameters

Let $0 < p < 1$, and let $\{X_1, \ldots, X_n\}$ be a random sample of observable R.V.'s that depend on some parameter θ.

- If T_1 and T_2 are statistics of $\{X_1, \ldots, X_n\}$ such that $\Pr(T_1 < \theta < T_2) \geq p$ for all possible values of θ, then we say that the random open interval (T_1, T_2) is a 100p percent confidence interval for θ.
- If T_1 and T_2 are statistics of $\{X_1, \ldots, X_n\}$ such that $\Pr(T_1 < \theta < T_2) = p$ for all possible values of θ, then the 100p percent confidence interval (T_1, T_2) is called exact.
- After the observed values $X_1 = x_1, \ldots, X_n = x_n$ are given, let $T_1 = t_1$ and $T_2 = t_2$ be the corresponding computed values. Then the open interval (t_1, t_2) is called the observed value of the confidence interval.

Important Note: A confidence interval is random! It is a **pair of R.V.'s** forming a random open interval.

- Different observed values for $X_1, \ldots, X_n$ give different actual open intervals.

Interpretation of Confidence Intervals

Example: Let $\{X_1, \ldots, X_n\}$ be a random sample of observable R.V.'s that depend on some parameter θ, and suppose that (T_1, T_2) is a 95% confidence interval for θ.

Interpretation:

- T_1, T_2 are statistics of $\{X_1, \ldots, X_n\}$.
- A confidence interval (T_1, T_2) is a random open interval. Different observed values for $X_1, \ldots, X_n$ give different observed values for the confidence interval (T_1, T_2).
- By saying that (T_1, T_2) is a 95% confidence interval for θ, it means that 95% of all observed values (t_1, t_2) for (T_1, T_2) are open intervals that actually contain θ.
  - It does **NOT** mean that every observed open interval (t_1, t_2) has a 95% probability of containing θ.
  - The "95%" relates to the entire estimation procedure, and not to a specific open interval.
  - 95% of all possible open intervals contain the parameter θ. Each specific (t_1, t_2) either contains θ, or doesn't contain θ.

Unbiased estimators

Let $X_1, \ldots, X_n$ be observable R.V.'s whose joint distribution is parametrized by some parameter θ with parameter space Ω.

- An estimator of θ is a real-valued function $\delta(X_1, \ldots, X_n)$.
- **Recall:** The sampling distribution of δ is the distribution of δ.
- This sampling distribution of δ depends on the parameter θ.
- For every possible value θ in the parameter space Ω, the mean of the sampling distribution of δ given $\theta = \theta$, is denoted by $\mathbf{E}_\theta[\delta(X_1, \ldots, X_n)]$.

Definition: An estimator $\delta(X_1, \ldots, X_n)$ is called an unbiased estimator of θ if $\mathbf{E}_\theta[\delta(X_1, \ldots, X_n)] = \theta$ for every possible value θ in the parameter space Ω.

Example: If $X_1, \ldots, X_n$ are normal R.V.'s that are conditionally iid given the mean μ, then the **sample mean** is an unbiased estimator of μ, since the mean of $\overline{X}_n$ given $\mu = \mu_0$ is precisely μ_0.

- If each X_i has mean 5, then the sample mean has mean 5; If each X_i has mean 2, then the sample mean has mean 2, etc.

Biased estimators and bias

Let $X_1, \ldots, X_n$ be observable R.V.'s whose joint distribution is parametrized by some parameter θ with parameter space Ω.

An estimator $\delta(X_1, \ldots, X_n)$ is called a biased estimator of θ if it is not an unbiased estimator of θ, or equivalently, if there is some θ in the parameter space Ω such that $\mathbf{E}_\theta[\delta(X_1, \ldots, X_n)] \neq \theta$.

- The bias of an estimator $\delta(X_1, \ldots, X_n)$ is a function defined on Ω, such that each $\theta \in \Omega$ is mapped to $\mathbf{E}_\theta[\delta(X_1, \ldots, X_n)] - \theta$.

Interpretation: Let $\delta = \delta(X_1, \ldots, X_n)$ be an estimator of some parameter θ with parameter space Ω. If for every possible value θ in Ω, the mean of the estimator is exactly θ, then the bias of δ is the zero function.

Why the unbiased sample variance is unbiased

Theorem: The unbiased sample variance is an unbiased estimator.

Proof: Recall that the unbiased sample variance of $\{X_1, \ldots, X_n\}$ is

$$s_n^2(X_1, \ldots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2.$$

Since $X_i - \overline{X}_n = (X_i - \mu) - (\overline{X}_n - \mu)$ for every i, we get

$$\begin{aligned}
\sum_{i=1}^{n} (X_i - \overline{X}_n)^2 &= \sum_{i=1}^{n} \left((X_i - \mu) - (\overline{X}_n - \mu)\right)^2; \\
&= \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X}_n - \mu) \sum_{i=1}^{n} (X_i - \mu) + n(\overline{X}_n - \mu)^2 \\
&= \sum_{i=1}^{n} (X_i - \mu)^2 - 2(\overline{X}_n - \mu) \cdot n(\overline{X}_n - \mu) + n(\overline{X}_n - \mu)^2 \\
&= \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X}_n - \mu)^2.
\end{aligned}$$

Why the unbiased sample variance is unbiased (continued)

Consequently,

$$\begin{aligned}
\mathbf{E}_{\sigma^2}[s_n^2(X_1,\ldots,X_n)] &= \mathbf{E}\Big[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\Big] \\
&= \mathbf{E}\Big[\tfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\tfrac{n}{n-1}(\overline{X}_n-\mu)^2\Big] \\
&= \mathbf{E}\Big[\tfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2\Big]-\mathbf{E}\big[\tfrac{n}{n-1}(\overline{X}_n-\mu)^2\Big] \\
&= \tfrac{n}{n-1}\mathbf{E}\Big[\tfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\Big]-\tfrac{n}{n-1}\mathbf{E}\big[(\overline{X}_n-\mu)^2\big] \\
&= \tfrac{n}{n-1}\cdot\sigma^2-\tfrac{n}{n-1}\cdot\tfrac{\sigma^2}{n} \\
&= \sigma^2.
\end{aligned}$$

Therefore $\mathbf{E}_{\sigma^2}[s_n^2(X_1,\ldots,X_n)]-\sigma^2$ is identically the zero function, which implies that the bias of $\delta(X_1,\ldots,X_n)$ is zero, i.e. δ is unbiased.

Intuition of unbiased estimation

Let $X_1, \ldots, X_n$ be observable R.V.'s whose joint distribution is parametrized by some parameter θ with parameter space Ω.

Let $\delta(X_1, \ldots, X_n)$ be an estimator of θ.

- Notice that $X_1, \ldots, X_n$ are parametrized by θ.
- For $\delta(X_1, \ldots, X_n)$ to be a "good" estimator of θ, we want a high probability that the estimator δ given some value θ is close to θ.
- Hence one useful property for such a "good" estimator is that the mean of δ given θ should equal θ, which is precisely what it means for δ to be unbiased.

Note: "unbiased estimator" is one of several possible ways to define a "good" estimator, but it is not the only way.

- e.g. the M.L.E. of the variance of a random sample of normal R.V.'s is the biased sample variance, which is also a "good" estimator of the variance.

Summary

- t-distribution
- t-distribution versus normal distribution
- Confidence interval
- Unbiased estimation