Stat210A: Theoretical Statistics **Lecture Date: September 18, 2014**

Lecture 7: t-Distribution and the Information Inequality

Lecturer: Michael I. Jordan *Scribe: Virginia Smith*

1 The t-Distribution

Let $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, $\bar{X} = \frac{X_1+\cdots+X_n}{n}$, and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Define:

$$Z \overset{\text{def}}{=} \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \tag{1}$$

$$V \overset{\text{def}}{=} \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \tag{2}$$

Recall from last lecture that $\bar{X}$ and S^2 are independent due to Basu's Theorem, and that the random variable V, along with $\bar{X} \sim N(\mu, \sigma^2/n)$, implicitly determines the joint distribution of $\bar{X}$ and S^2. The variables Z and V are called *pivots*, as their distributions do not depend on the unknown parameters μ and σ^2. Note that they are not, however, statistics, as they both depend explicitly on unknown parameters. Z and V are independent because they are functions of $\bar{X}$ and S^2, respectively, and so the following will also be a pivot:

$$T = \frac{Z}{\sqrt{V/(n-1)}} \tag{3}$$

$$= \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} \tag{4}$$

$$= \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{5}$$

This variable is distributed according to the t-distribution, which we define below.

Definition 1 (The t-Distribution)**.** Suppose we have a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$. Let $\bar{X} = \frac{X_1+\cdots+X_n}{n}$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. The resulting t-value:

$$T \overset{\text{def}}{=} \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{6}$$

is distributed according to the t-distribution with $n-1$ degrees of freedom, denoted $\sim t_{n-1}$. The density is:

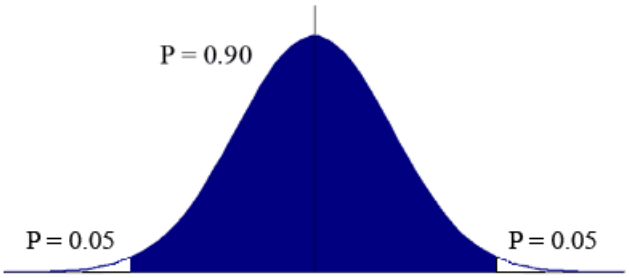
$$f_T(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)(1 + x^2/\nu)^{(\nu+1)/2}}, \tag{7}$$

for $x \in \mathbb{R}$ and where $\nu = n - 1$ denotes the number of degrees of freedom. The t-distribution is bell-shaped and symmetric like the normal distribution, but with heavier tails.

Pivots like the t-value are widely used in (1) developing confidence intervals and (2) hypothesis testing. We briefly discuss both below.

1.1 Confidence Intervals

Suppose we choose a number $\kappa_{.95}$ such that $P(-\kappa_{.95} \leq T \leq \kappa_{.95}) = 0.90$, as depicted below[1]:



We can use $\kappa_{.95}$ to find a 90% confidence interval for μ, as the following two are equivalent:

$$P(-\kappa_{.95} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \kappa_{.95}) = 0.90$$
$$P(\bar{X} - \kappa_{.95}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + \kappa_{.95}\frac{S}{\sqrt{n}}) = 0.90$$

The interval with endpoints $\bar{X} \pm \kappa_{.05}\frac{S}{\sqrt{n}}$ is a 90% confidence interval for μ. Note that this has a somewhat unintuitive interpretation: μ is fixed, and the interval itself is random. The interpretation is that over many datasets, the random interval is expected to include μ 90% of the time.

1.2 Hypothesis Testing

The t-value can also be used for hypothesis testing. In the case that the true $\mu = 0$, we can directly use the value $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ to test the null hypothesis.

In the case that μ is not 0, we derive the noncentral t-distribution:

$$T = \frac{\bar{X}}{S/\sqrt{n}} \tag{8}$$
$$= \left(\frac{\bar{X} + \mu - \mu}{S/\sqrt{n}}\right)\frac{1/\sigma}{1/\sigma} \tag{9}$$
$$= \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} + \frac{\mu}{\sigma/\sqrt{n}}\right)\frac{1}{\sqrt{S^2/\sigma^2}} \tag{10}$$
$$= \frac{Z + \mu/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}\left(\frac{\mu}{\sigma/\sqrt{n}}\right) \tag{11}$$

Here T is said to be a noncentral t-distributed random variable with $n-1$ degrees of freedom and noncentrality parameter $\frac{\mu}{\sigma/\sqrt{n}}$. We will calculate the mean and variance of T in the section below, and will cover hypothesis testing in detail later in the course.

[1] http://www.chem.utoronto.ca/coursenotes/analsci/StatsTutorial/12tailed.html

1.3 The t-Distribution and Normal One-Sample Estimation

Before calculating the mean and variance of the t-distribution, let's visit a few important UMVU estimators for normal random variables.

Example 2 (σ^r)**.** To start, suppose that we wish to calculate an unbiased estimator for σ^r. Consider the following:

$$E[S^r] = E\left[\frac{\sigma^r}{(n-1)^{r/2}} V^{r/2}\right] \tag{12}$$

$$= \frac{\sigma^r}{(n-1)^{r/2}} \int_0^\infty \frac{x^{(r+n-3)/2} e^{-x/2}}{2^{(n-1)/2}\Gamma[(n-1)/2]} dx \tag{13}$$

$$= \frac{\sigma^r 2^{r/2}\Gamma[(r+n-1)/2]}{(n-1)^{r/2}\Gamma[(n-1)/2]} \tag{14}$$

Therefore, the following will be an unbiased estimator of σ^r:

$$\frac{(n-1)^{r/2}\Gamma[(n-1)/2]}{2^{r/2}\Gamma[(r+n-1)/2]} S^r \tag{15}$$

Further, this estimate will be UMVU because it is a function of the complete sufficient statistic $(\bar{X}, S^2)$. When $r = 2$, this estimate reduces to the familiar UMVU estimate S^2. Also note that, though it is biased, the natural choice of using S^r to estimate σ^r will have only slight bias as $n \to \infty$. Using Stirling's formula:

$$\frac{(n-1)^{r/2}\Gamma[(n-1)/2]}{2^{r/2}\Gamma[(r+n-1)/2]} = 1 - \frac{r(r-2)}{4n} + O(1/n^2) \tag{16}$$

Example 3 (μ^2)**.** Recall that $\bar{X}$ is UMVU for μ. However, $\bar{X}^2$ will be a biased estimator, since:

$$E[\bar{X}^2] = \mathrm{Var}(\bar{X}) + E[\bar{X}]^2 \tag{17}$$

$$= \sigma^2/n + \mu^2 \tag{18}$$

This bias can be removed by subtracting S^2/n, which we know is an unbiased estimate of σ^2/n. Thus $\bar{X}^2 - S^2/n$ is UMVU for μ^2.

Example 4 (μ/σ)**.** The quantity μ/σ is known as the signal-to-noise ratio. We already know that $\bar{X}$ is unbiased for μ. We can find an unbiased estimator for σ by using the formula derived in Example 2. This estimate is independent of $\bar{X}$, and so we can multiply them together to find:

$$\frac{\bar{X}\sqrt{2}\Gamma[(n-1)/2]}{S\sqrt{n-1}\Gamma[(n-2)/2]} \tag{19}$$

is UMVU for μ/σ.

Using these results, we can easily calculate the mean and variance of the noncentral t-distribution. Independence follows from Basu's theorem, giving us:

$$E[T] = \sqrt{n}E[\bar{X}/S] = \sqrt{n}E[\bar{X}]E[S^{-1}] \tag{20}$$

Plugging in results from Example 2:

$$E[T] = \left(\frac{\mu}{\sigma/\sqrt{n}}\right)\left(\frac{\sqrt{n-1}\Gamma[(n-2)/2]}{\sqrt{2}\Gamma[(n-1)/2]}\right) \tag{21}$$

We can also calculate the variance:

$$\begin{aligned}
\text{Var}(T) &= E[T^2] - E[T]^2 && (22)\\
&= nE[\bar{X}^2/S^2] - E[T]^2 && (23)\\
&= nE[\bar{X}^2]E[S^2] - E[T]^2 && (24)\\
&= n\left(\sigma^2/n + \mu^2\right)\left(\sigma^{-2}\frac{n-1}{n-2}\right) - E[T]^2 && (25)\\
&= (1+\delta^2)\frac{n-1}{n-3} - E[T]^2 && (26)\\
&= \frac{n-1}{n-3} + \delta^2\left(\frac{n-1}{n-3} - \frac{(n-1)\Gamma^2[(n-2)/2]}{2\Gamma^2[(n-1)/2]}\right) && (27)
\end{aligned}$$

Where $\delta = \frac{\mu}{\sigma/\sqrt{n}}$, using results from the previous examples.

2 The Information Inequality

Theorem 5 (Information Inequality, Thm 4.9 in Keener)**.** *Let $P = \{P_\theta : \theta \in \Omega\}$ be a dominated family with Ω an open set in $\mathbb{R}$ and densities p_θ differentiable with respect to θ. If $E_\theta[\psi] = 0$, $E_\theta[\delta^2] < \infty$, and $g'(\theta) = E_\theta[\delta\psi]$ or $g'(\theta) = \int \delta\psi p_\theta d\mu$ hold for all $\theta \in \Omega$, then:*

$$Var_\theta(\delta) \geq \frac{[g'(\theta)]^2}{I(\theta)}, \qquad \theta \in \Omega \tag{28}$$

This result is called the information, or Cramér-Rao, inequality. The last regularity condition can be somewhat troublesome. One way to deal with this condition is to impose more restrictive conditions on the model P, and show that the bound holds for all δ at all $\theta \in \Omega$.

The term I is known as the Fisher information, given by:

$$I(\theta) = E_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta}\right)^2 \tag{29}$$

Under regularity conditions, we also have:

$$I(\theta) = -E_\theta\frac{\partial^2 \log p_\theta(X)}{\partial\theta^2} \tag{30}$$

This gives us a nice interpretation of the Fisher information as the curvature of the problem. We can see that if this curvature is low, the information inequality bound will be larger, meaning that it could be more difficult to find the best value for θ.

We begin our derivation of the information inequality by bounding the variance, through an application of Cauchy-Schwarz.

Theorem 6 (Cauchy-Schwarz)**.** *For all vectors x and y in an inner product space:*

$$\langle x, y\rangle \leq ||x|| \cdot ||y|| \tag{31}$$

We can apply this to random variables by noting we can define an inner product space, consisting of real-valued random variables on a fixed probability space. The inner product $\langle X, Y \rangle = E[XY]$ preserves the notion of an inner product in this space. Using $\text{Cov}(X, Y) = \langle X - \mu_X, Y - \mu_Y \rangle$ gives us the *covariance inequality*:

Lemma 7 (Covariance Inequality)**.** *Let X and Y be random variables. Then,*

$$Cov^2(X, Y) \leq Var(X)\, Var(Y) \tag{32}$$

Using this inequality with δ and an unbiased estimator of $g(\theta)$ and arbitrary random variable ψ, we have:

$$\text{Var}_\theta(\delta) \geq \frac{\text{Cov}_\theta^2(\delta, \psi)}{\text{Var}_\theta(\psi)} \tag{33}$$

An immediate issue is that the right hand side of this inequality involves δ, making it a somewhat useless bound for the variance of δ. We will see next lecture that if ψ is chosen cleverly we can remedy this problem, finding a $\text{Cov}_\theta(\delta, \psi)$ that will remain the same for all δ unbiased for $g(\theta)$.

References

Keener, R. (2010). *Theoretical Statistics: Topics for a Core Course.* Springer, New York, NY.