

Econometrics – Inference

William M Volckmann II

January 3, 2017

1 Introduction

Sometimes we'll have a random variable X but its exact distribution function is not known. In particular, we might know the general form of $f(x)$ but it will contain some unknown parameter θ . For example, we might know that X has an exponential distribution so that

$$f(x) = \theta e^{-\theta x},$$

but θ is unknown. We often denote this problem by saying that X has a density $f(x; \theta)$. In this case, θ is called the **parameter** of the distribution. We want to estimate θ .

Our information about θ will come from a sample on X . The sample observations have the same distribution as X , and we will denote them as the random variables X_1, X_2, \dots, X_n , where n denotes the **sample size**. When the sample is drawn, we use x_1, x_2, \dots, x_n as the values of the **realizations** of the sample. We will often make the following assumption about the samples.

Definition 1. If the random variables X_1, X_2, \dots, X_n are independent and identically distributed (iid), then these random variables constitute a **random sample** of size n from the common distribution.

Much of the time we will use functions of the sample to summarize the information contained within the sample. These are called statistics.

Definition 2. Let X_1, \dots, X_n denote a sample on a random variable X . Let $T = (X_1, \dots, X_n)$ be a function of the sample. Then T is called a **statistic**.

Once the sample is drawn, t is called the realization of T , that is, $t = T(x_1, \dots, x_n)$. We will be considered a statistic T that functions as an **estimator** of θ . (More formally, a point estimator of θ .) The realization of T is called an **estimate** of θ .

Definition 3. Let X_1, \dots, X_n denote a sample on a random variable X with pdf $f(x; \theta)$. Let $T = T(X_1, \dots, X_n)$ be a statistic. We say that T is an **unbiased** estimator of θ if $E[T] = \theta$.

2 Maximum Likelihood Estimator

Suppose X_1, \dots, X_n are form a random sample, that is, are iid. Then the joint distribution of the random sample is

$$\prod_{i=1}^n f(x_i; \theta).$$

We want to view this as a function of θ , so let's write it as

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta).$$

The function $L(\theta)$ is called the **likelihood function** of the random sample. An common estimate of θ is the value of θ that maximizes $L(\theta)$. If there is a unique estimate satisfying

$$\hat{\theta} = \arg \max L(\theta),$$

then we call $\hat{\theta}$ the **maximum likelihood estimator (MLE)**.

It is often easier to work with the log of the likelihood function, so let's

have $l(\theta) = \ln(L(\theta))$. Since a logarithm is a monotonically increasing function, the maximizer is unchanged by such a transformation and thus we acquire the same MLE. Many models are differentiable, so $\hat{\theta}$ often solves

$$\frac{\partial l(\theta)}{\partial \theta} = 0.$$

3 Student's Theorem

And now for a brief aside.

Theorem 1. *Let X_1, \dots, X_n be iid random variables, each having a normal distribution with mean μ and variance σ^2 . Define the random variables*

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

Then

- (a)** \bar{X} has a $\mathcal{N}(\mu, \sigma^2/n)$ distribution.
- (b)** \bar{X} and S^2 are independent.
- (c)** $(n-1)S^2/\sigma^2$ has a $\chi^2(n-1)$ distribution.
- (d)** The random variable

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

has a Student t -distribution with $n-1$ degrees of freedom.

4 Confidence Intervals

Suppose L and U are two statistics for $f(x; \theta)$. We say that the interval (L, U) is a $(1 - \alpha)100\%$ **confidence interval** for θ if

$$P[\theta \in (L, U)] = 1 - \alpha.$$

Suppose $\alpha = .05$. Then $P[\theta \in (L, U)] = 0.95$ is a 95% confidence interval. In this case, $1 - \alpha = 0.95$ is the **confidence coefficient** of the interval.

Once the sample is drawn, the realized value of the confidence interval is (l, u) . Either the interval contains θ or it doesn't. The idea is that 95% of our realized confidence intervals will contain θ , so we can feel 95% confident that θ lies in any given realized interval (l, u) .

One measure of efficiency based on confidence intervals is their expected length. Suppose (L_1, U_2) and (L_2, U_2) have the same confidence coefficient. Then we can say that (L_1, U_1) is more efficient than (L_2, U_2) if $E_\theta[U_1 - L_1] \leq E_\theta[U_2 - L_2]$.

4.1 Confidence Interval for μ Under Normality

Suppose X_1, \dots, X_n are random samples from $\mathcal{N}(\mu, \sigma^2)$. Let \bar{X} and S^2 denote the sample mean and sample variance, respectively. The random variable

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

has a t distribution with $n - 1$ degrees of freedom.

For $0 < \alpha < 1$, define $t_{\alpha/2}^{n-1}$ to be the upper $\alpha/2$ critical point of a t distribution with $n - 1$ degrees of freedoms. Which is to say, $P(T > t_{\alpha/2}^{n-1}) = \alpha/2$, and similarly, $P(T < -t_{\alpha/2}^{n-1}) = \alpha/2$. Therefore, we can write

$$P\left([T < -t_{\alpha/2}^{n-1}] \text{ and } [T > t_{\alpha/2}^{n-1}]\right) = \alpha.$$

It follows that

$$P(-t_{\alpha/2}^{n-1} \leq T \leq t_{\alpha/2}^{n-1}) = 1 - \alpha.$$

Now just do some algebra to get

$$\begin{aligned}
1 - \alpha &= P\left(-t_{\alpha/2}^{n-1} \leq T \leq t_{\alpha/2}^{n-1}\right) \\
&= P\left(-t_{\alpha/2}^{n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{\alpha/2}^{n-1}\right) \\
&= P\left(\bar{X} - t_{\alpha/2}^{n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}^{n-1} \frac{S}{\sqrt{n}}\right).
\end{aligned}$$

Once the sample is drawn, let \bar{x} and s denote the realized values of \bar{X} and S . Then a $(1 - \alpha)100\%$ confidence interval for μ is given by

$$\left(\bar{x} - t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}}\right).$$

This interval is often referred to as the $(1 - \alpha)100\%$ **t-interval** for μ . The estimate of the standard deviation of \bar{X} , s/\sqrt{n} , is called the **standard error** of \bar{X} .

Note that if σ is known, then we can instead use the statistic

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

which is standard normal distributed.

4.2 Confidence Intervals for μ Without Normality

The confidence intervals for normally distributed random variables is exactly true. If the samples are not normally distributed, however, then it might still be approximately true. In particular, the **Central Limit Theorem** shows that the distribution of T is approximately standard normal for large enough n .

Theorem 2 (Central Limit Theorem). *Let X_1, \dots, X_n denote the observations of a random sample from a distribution that has mean μ and finite*

variance σ^2 . Then the distribution function of the random variable

$$W_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

converges to the distribution function of the standard normal distribution, as $n \rightarrow \infty$.

The result stays the same if we replace σ with the sample standard deviation S , in which case we have

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

approximately distributed according to standard normal. So for “large n ,” we’ll have

$$\begin{aligned} 1 - \alpha &\approx P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < z_{\alpha/2}\right) \\ &= P\left(\bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}\right). \end{aligned}$$

Thus, given realizations \bar{x} and s , the approximate $(1 - \alpha)100\%$ **large sample confidence interval** for μ is

$$\left(\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}\right).$$

In practice we usually don’t know if the population is normally distributed. But for the most part, given the same α , the intervals using the t -distribution are larger than those using central limit theorem. To be conservative, statisticians usually prefer the t -distribution-based confidence intervals.

5 Hypothesis Testing

Consider $f(x; \theta)$, where Ω is the set of values θ could possibly take on. Suppose we believe that $\theta \in \omega_0 \subseteq \Omega$. Define $\omega_1 = \Omega \setminus \omega_0$. Then our **null hypothesis**, denoted H_0 , is $H_0 : \theta \in \omega_0$. We want to test the null hypothesis against the **alternative hypothesis**, $H_1 : \theta \in \omega_1$.

Usually we'll want the null hypothesis to be something along the lines of "no effect" or "no change." The alternative is often referred to as the research worker's hypothesis. One-sided and two-sided tests are common. In a **two-sided** test, we will test, for example, $H_0 : \theta = 0$ against $H_a : \theta \neq 0$. In this case, we would *reject* the null hypothesis if our estimate of θ was too big in either the positive or negative direction. In a **one-sided** test, we might instead test $H_0 : \theta \leq 0$ against $H_a : \theta > 0$. In this case, we would only reject the null hypothesis if our estimate of θ was too big in the positive direction.

We will define some **critical region** such that if our realization of the random variables fall within this region, we *reject the null hypothesis*. If our realization of the random variables does not fall within the critical region, then we *fail to reject the null hypothesis*. Notice that we are not *confirming* either hypothesis – we are simply making a **test** about whether to reject the null or not.

The decision to reject or not reject H_0 is based on a sample X_1, \dots, X_n from the distribution of X and thus, the decision could be wrong. Indeed, there are two types of errors we could make. A **type 1 error** occurs when we reject the null hypothesis even though it is true. A **type 2 error** occurs when we fail to reject the null hypothesis even though it is false.

Unfortunately, we cannot simultaneously minimize the probability of both types of errors. For instance, the probability of a type 1 error can only be minimized if we never reject the null hypothesis – and this, of course, means we'll be committing a type 2 error quite a bit since we'll have no way of filtering out false null hypotheses. The solution is to select a fixed probability

of committing a type 1 error, and then from there minimize the probability of committing a type 2 error.

To that end, we define the **size** of a test to be the probability of rejecting a true null hypothesis, i.e. the probability of committing a type 1 error. A critical region has size α if

$$\alpha = \max_{\theta \in \omega_0} P[(X_1, \dots, X_n) \in C].$$

The **power** of a test is the probability to reject a false null hypothesis, or equivalently, $1 - P(\text{type 2 error})$.

5.1 Large Sample Test for Mean (One-Sided)

Suppose X is a random variable with mean μ and variance σ^2 . We want to test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$. Since \bar{X} is unbiased for μ , an obvious test is to reject H_0 if \bar{X} is much larger than μ_0 . The question is, how do we define “much larger than?”

Using the Central Limit Theorem, we know that

$$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Thus, we reject H_0 if $Z \geq z_\alpha$ for size α . Alternatively, we could write the rejection rule as

$$\bar{X} \geq z_\alpha \frac{s}{\sqrt{n}} + \mu_0. \tag{1}$$

Now let’s calculate the power function. We start by asking, what is the probability that we reject the null hypothesis, given that the null hypothesis

is false? In other words, what is the probability that equation (1) is satisfied?

$$\begin{aligned} P\left(\bar{X} \geq z_\alpha \frac{s}{\sqrt{n}} + \mu_0\right) &= P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \geq \frac{z_\alpha \frac{s}{\sqrt{n}}}{s/\sqrt{n}} + \frac{\mu_0}{s/\sqrt{n}} - \frac{\mu}{s/\sqrt{n}}\right) \\ &= P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \geq z_\alpha + \frac{\mu_0 - \mu}{s/\sqrt{n}}\right). \end{aligned}$$

At this point we can replace s with σ if our sample size is large enough. Then it is approximately true that

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

so we can write the probability as

$$\begin{aligned} P\left(Z \geq z_\alpha + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) &= 1 - P\left(Z \leq z_\alpha + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(z_\alpha + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right). \end{aligned}$$

Such is the power function. Notice that as μ increases, the power function approaches one. This is intuitive: the large μ actually is, the more likely we're actually going to reject the null that $\mu = \mu_0$.

Of course, we did cheat a little here, kind of. We appealed to the Central Limit Theorem when we replaced s with σ . Ideally we would want to use the t -statistic with $n - 1$ degrees of freedom when σ^2 is unknown. So what we've done here holds for *very* large n . However, if n is only *kinda-sorta* large¹, then use the t -distribution instead.

5.2 Large Sample Test for Mean (Two-Sided)

I won't go knee-deep here, but I'll briefly comment on the power function. When we're dealing with a two-sided test, we will have another component

¹This is a technical term.

to the power function. This is because in a two-sided test, we reject the null if our statistic is either $Z \geq z_{\alpha/2}$ or $Z \leq -z_{\alpha/2}$.

The probability that our statistic is too high is, as you might expect,

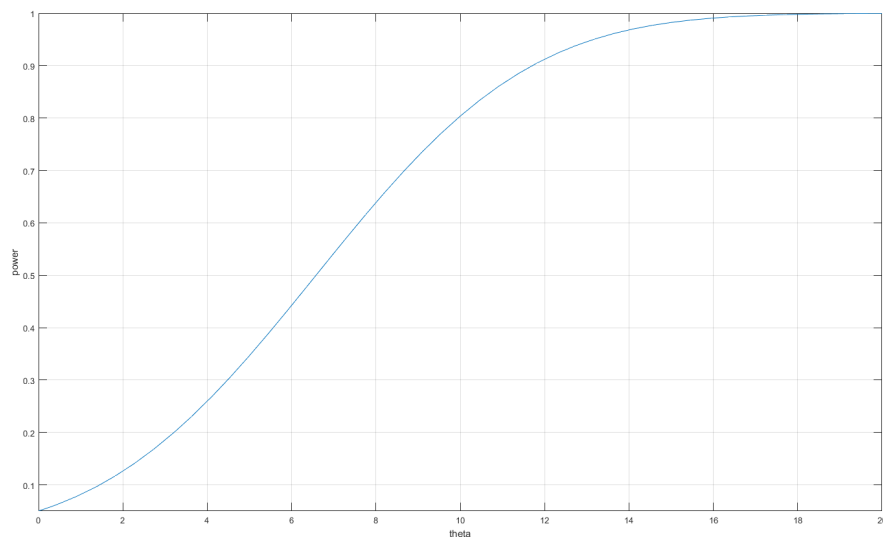
$$1 - \Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right).$$

The probability that our statistic is too low, on the other hand, is

$$\begin{aligned} P\left(\bar{X} \leq -z_{\alpha/2} \frac{s}{\sqrt{n}} + \mu_0\right) &= P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \leq \frac{-z_{\alpha/2} \frac{s}{\sqrt{n}}}{s/\sqrt{n}} + \frac{\mu_0}{s/\sqrt{n}} - \frac{\mu}{s/\sqrt{n}}\right) \\ &\approx P\left(Z \leq -z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right). \end{aligned}$$

Therefore, the two-sided power function is

$$\gamma_C(\mu) = 1 - \Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right).$$



The shape of a typical power function.