Go home p-value, you're drunk

Jose Luis Avilez

2018-07-24

In this lecture we strive to prove a pseudo-theorem by C. Shalizi:

Theorem 0.1. ***Pseudo-theorem (Shalizi)*** *Any p-value distinguishable from zero is insufficiently informative.*

This proposition is crucial for two reasons:

1. First, it ride many Pure Mathematics students from studying quite a useless topic during STAT 231/241 and instead allows them to immerse themselves in probability theory.

2. It saves us from the pain of seeing many users of statistics say things which are not true.

For the purposes of this talk, we assume that everyone is familiar with what a random variable is (an informal understanding is fine) and what a convergent sequence is. For a few preliminaries, we extend our notion of convergence to random variables.

0.1 Preliminaries

Definition 0.2. A sequence of random variables (X_n) with cumulative density functions $F_n(x)$ is said to **converge in distribution** to X if

$$\lim_{n\to\infty} F_n(x) = F(x)$$

where $F(x)$ is the cumulative distribution of X. We write $X_n \xrightarrow{D} X$.

Definition 0.3. A sequence of random variables (X_n) with cumulative density functions $F_n(x)$ is said to **converge in probability** to X if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P[|X_n - X| \geq \epsilon] = 0$$

or equivalently

$$\lim_{n\to\infty} P[|X_n - X| < \epsilon] = 0$$

and we write $X_n \xrightarrow{P} X$

Theorem 0.4. *If X is a random variable and $u(X)$ is a nonnegative real-values function such that $\mathrm{E}[u(X)]$ exists, then for any positive constant $c > 0$,*

$$P[u(X) \geq c] \leq \frac{\mathrm{E}[u(X)]}{c}$$

Proof. We argue as follows:

$$\begin{aligned}
\mathrm{E}[u(X)] &= \int_{x\in A} u(x)f(x)dx + \int_{x\notin A} u(x)f(x) \\
&\geq \int_{x\in A} u(x)f(x)dx \\
&\geq \int_{x\in A} cf(x)dx \\
&= c\int_{x\in A} f(x)dx \\
&= cP(X \in A) \\
&= cP(u(X) \geq c)
\end{aligned}$$

Which completes the proof. ■

Theorem 0.5. ***Chebyshev's Inequality.*** *Suppose X is a random variable with a finite mean μ and finite variance σ^2. Then for any $k > 0$,*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

or, equivalently,

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Proof. Let $u(X) = (X - \mu)^2$ and $c = k^2\sigma^2$. Thus, by Theorem 0.4, we have

$$P[(X - \mu)^2 \geq k^2\sigma^2] \leq \frac{\mathrm{E}\left[(X - \mu)^2\right]}{k^2\sigma^2}$$

But $\mathrm{E}\left[(X - \mu)^2\right] = \sigma^2$ and we can take the square root inside the argument of the probability, thus yielding

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

■

Theorem 0.6. ***Weak Law of Large Numbers.*** *Let (X_n) be a sequence of independent and identically distributed random variables with common mean μ and finite variance σ^2. Define the random variable $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then,*

$$\overline{X}_n \xrightarrow{P} \mu$$

Proof. We use Chebyshev's Inequality. Note that the mean and variance of $\overline{X}_n$ are μ and $\frac{\sigma^2}{n}$, respectively. Fix $\epsilon > 0$. Then,

$$\begin{aligned} P[|\overline{X}_n - \mu| \geq \epsilon] &= P\left[|\overline{X}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}\right] \qquad \text{where } k = \frac{\epsilon\sqrt{n}}{\sigma} \\ &\leq \frac{\sigma^2}{n\epsilon^2} \to 0 \end{aligned}$$

as required. ■

Theorem 0.7. ***Central Limit Theorem.*** *Suppose $X_1, X_2, \ldots$ is a sequence of independent random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then*

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \xrightarrow{D} N(0, 1)$$

Proof sketch. Assume that each of these has a moment generating function. Look at the Taylor series expansion of the moment generating function for $X_i - \mu$. Use the information provided and show that the moment generating functions converge to the moment generating function of a standard normal random variable. ■

0.2 Shalizi's Theorem

We now need to define what a p-value is. We shall see why such a definition should never be present in a Statistics course.

Definition 0.8. A **null hypothesis** is the default position under the model assumptions and is usually denoted as H_0. Opposite to it, we have the **alternative hypothesis**.

Definition 0.9. A **discrepancy measure** or a **test statistic** is a measure that evaluates if the data supports H_0. We want to test if the discrepancy measure is extreme enough to reject the null hypothesis.

Definition 0.10. Let $X_1, \ldots, X_n$ is a sample and suppose that $f(X; \theta; H_0)$ is the distribution under the null hypothesis. Suppose D is a discrepancy measure used to test the hypothesis. The **p-value** of the hypothesis test is defined as

$$p = P(D \geq |X| \mid H_0)$$

Theorem 0.11. *Let $X_1, \ldots, X_n$ be a sample used to estimate the true mean μ of a population. Then, the p-value of the hypothesis test using the estimator $\hat{\mu}$ is useless.*

Proof. For simplicity, let's assume we want to test whether the mean is zero or not. Under the null hypothesis, the sampling distribution of the mean is $N(\mu, \frac{\sigma^2}{n})$. From this, our test statistic will be $\frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}}$.

By the CLT, $\frac{\sqrt{n}}{\sigma^2}(\hat{\mu} - \mu) \xrightarrow{D} N(0,1)$. Re-arranging this equation, we obtain $\hat{\mu} \xrightarrow{D} \mu + \frac{\sigma}{\sqrt{n}} N(0,1)$. We can write the estimator as the sum of a deterministic and stochastic component. That is, $\hat{\mu} = \mu + O(n^{-\frac{1}{2}})$. We recall that

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

so that we can write $\hat{\sigma} = \sigma + O(n^{-\frac{1}{2}})$. Thus, we can write our test statistic as

$$\begin{aligned}
\frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}} &= \sqrt{n}\frac{\hat{\mu}}{\hat{\sigma}} \\
&= \sqrt{n}\frac{\mu + O(n^{-\frac{1}{2}})}{\sigma + O(n^{-\frac{1}{2}})} \\
&= \sqrt{n}\left(\frac{\mu}{\sigma} + O(n^{-\frac{1}{2}})\right) \\
&= \sqrt{n}\frac{\mu}{sigma} + O(1)
\end{aligned}$$

Thus, as n grows, our test statistic grows to either ∞ or $-\infty$ unless the mean is precisely zero. Thus, the p-value tends to zero as $n \to \infty$. ∎

Theorem 0.12. *In a simple linear regression model, the p-value to test whether $\beta \neq 0$ is meaningless.*

Proof. We derive $\hat{\beta} = (X^T X)^{-1} X^T y$ and $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. Note that the variance covariance matrix of X is $\frac{1}{n} X^T X = V$. Thus,

$$\text{Var}\left(\hat{\beta}\right) = \frac{\sigma^2}{n} V^{-1}$$

This shows that the variance of the test statistic will vary at the rate of $\frac{1}{n}$, making p vanish whenever $\beta \neq 0$. ∎

Definition 0.13. We say an estimator $\hat{\theta}$ is consistent if

$$\hat{\theta} \xrightarrow{P} \theta$$

This is equivalent to saying that the probability of Type I and Type II errors converge to zero. It is easy to see that whenever we have a sequence of consistent hypothesis tests, the p-value converges stochastically to zero. We see this because we have an interval $[0, a_n]$ over which we reject the null hypothesis and an interval $(a_n, 1]$ on which we do not reject the null hypothesis. This can be formalised, but it is better to spend time discussing how fast does p converge to zero.

Theorem 0.14. *An upper bound to the tail probability of a normal distribution is*

$$P(Z \geq t) < \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-\frac{t^2}{2}}$$

A lower bound for the tail probability is

$$\frac{\exp(-t^2/2)}{(t^2+1)\sqrt{2\pi}}$$

Proof. Take

$$P(Z \geq t) = \int_t^\infty f(t)dx < \int_t^\infty \frac{x}{t} f(t)dx$$

and the result follows. ∎

Theorem 0.15. *The p-values for tests of means converge stochastically to zero at an exponential rate.*

Proof. By the Lemma above, we have

$$P_n = P\left(|Z| > \left|\frac{\hat{\mu}}{\hat{\sigma}/\sqrt{n}}\right|\right) \leq 2\frac{\exp(-n\hat{\mu}/(2\hat{\sigma}^2))}{\sqrt{n}\hat{\mu}\sqrt{2\pi}/\hat{\sigma}}$$

Taking the logarithm of both sides and dividing by n, we obtain,

$$\frac{1}{n}\log P_n \leq \frac{\log 2}{n} - \frac{\hat{\mu}^2}{2\hat{\sigma}^2} - \frac{\log n}{2n} - \frac{1}{n}\log\frac{\hat{\mu}}{\hat{\sigma}} - \frac{\log 2\pi}{n}$$

So that

$$\lim_{n\to\infty}\frac{1}{n}\log P_n \leq -\frac{\mu^2}{2\sigma^2}$$

Using the lemma above, and arguing similarly,

$$\lim_{n\to\infty}\frac{1}{n}\log P_n \geq -\frac{1}{2}\frac{\mu^2}{\sigma^2}$$

∎

The arguments above complete the proof of Shalizi's Theorem. We may ask, what is the use of p-values?