Theoretical Statistics. Lecture 6.

Peter Bartlett

1. U-statistics.
2. Projections.

Review. U-statistics

Definition: A U**-statistic of order** r **with kernel** h is

$$U = \frac{1}{\binom{n}{r}} \sum_{i \subseteq [n]} h(X_{i_1}, \ldots, X_{i_r}),$$

where h is symmetric in its arguments.

Review. U-statistics: Examples

- s_n^2 is a U-statistic of order 2 with kernel $h(x, y) = (1/2)(x - y)^2$.
- Kendall's τ: test for independence.
- Wilcoxon one-sample rank statistic: test for symmetry. Sum of U-statistics.

Review. Properties of U-statistics

- "U" for "unbiased": $\mathbf{E}U = \mathbf{E}h(X_1, \ldots, X_r)$.
- $\mathrm{Var}(U(X_1, \ldots, X_n)) \le \mathrm{Var}(h(X_1, \ldots, X_r))$ (Rao-Blackwell theorem).
- Concentration: If $|h(X_1, X_2)| \le B$ a.s., then

$$P(|U - \mathbf{E}U| \ge t) \le 2\exp(-nt^2/(8B^2)).$$

Review. Variance of U-statistics

$$\begin{aligned}
\mathrm{Var}(U) &= \frac{1}{\binom{n}{r}} \sum_{c=1}^{r} \binom{r}{c} \binom{n-r}{r-c} \zeta_c \\
&= \sum_{c=1}^{r} \theta(n^{-c}) \zeta_c, \\
\zeta_c &= \mathrm{Cov}(h(X_S), h(X_{S'})) \qquad \text{where } |S \cap S'| = c \\
&= \mathrm{Var}\left(\mathbf{E}\left[\, h(X_1^r) \middle|\, X_1^c\right]\right).
\end{aligned}$$

So if $\zeta_1 \neq 0$, the first term dominates:

$$n\mathrm{Var}(U) \to \frac{nr!(n-r)!r(n-r)!}{n!(r-1)!(n-2r+1)!}\zeta_1 \to r^2\zeta_1.$$

Variance of U-statistics: Example

Estimator of variance: $h(X_1, X_2) = (1/2)(X_1 - X_2)^2$:

$$\begin{aligned}
\zeta_1 &= \mathrm{Cov}(h(X_1, X_2), h(X_1, X_3)) \\
&= \mathrm{Var}(\mathbf{E}[h(X_1, X_2)|X_1]) + \mathbf{E}[\mathrm{Cov}(h(X_1, X_2), h(X_1, X_3)|X_1] \\
&= \mathrm{Var}(\mathbf{E}[h(X_1, X_2)|X_1]) = \mathrm{Var}\left(\mathbf{E}\left[\frac{1}{2}(X_1 - X_2)^2 | X_1\right]\right) \\
&= \mathrm{Var}\left(\mathbf{E}\left[\frac{1}{2}(X_1 - \mu + \mu - X_2)^2 | X_1\right]\right) \\
&= \mathrm{Var}\left(\frac{1}{2}((X_1 - \mu)^2 + \sigma^2)\right) = \frac{1}{4}(\mu_4 - \sigma^4),
\end{aligned}$$

where $\mu_4 = \mathbf{E}((X_1 - \mu)^4))$ is the 4th central moment. So $n\mathrm{Var}(U) \to \mu_4 - \sigma^4$.

We'll see that $\sqrt{n}(U - \sigma^2) \rightsquigarrow N(0, \mu_4 - \sigma^4)$. (What if $\mu_4 - \sigma^4 = 0$?)

Variance of U-statistics: Example

Recall Kendall's τ: For a random pair $P_1 = (X_1, Y_1), P_2 = (X_2, Y_2)$ of points in the plane, if X, Y are *independent and continuous* [and P_1P_2 is the line from P_1 to P_2]

$$h(P_1, P_2) = \left(1[P_1P_2 \text{ has positive slope}] - 1[P_1P_2 \text{ has negative slope}]\right),$$

$$\zeta_1 = \text{Cov}(h(P_1, P_2), h(P_1, P_3))$$

$$... = 1/9$$

so $n\text{Var}(U) \to 4/9$. We'll see that $\sqrt{n}U \rightsquigarrow N(0, 4/9)$. And this gives a test for independence.

Asymptotic distribution of U-statistics

How do we find the asymptotic distribution of a U-statistic?

We'll appeal to this theorem:

Theorem:

$$X_n \rightsquigarrow X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \Longrightarrow Y_n \rightsquigarrow X.$$

In particular, we find another sequence $\hat{U}$ such that

- $\sqrt{n}(U - \theta - \hat{U}) \xrightarrow{P} 0$, and
- The asymptotics of $\hat{U}$ are easy to understand.

In this case, we find $\hat{U}$ of the form $\hat{U} = \sum_i f(X_i)$. Then the CLT gives the result.

Asymptotic distribution of U-statistics

1. Why do functions of a single variable suffice? Because the interactions are weak.

2. How do we find suitable functions? By **projecting**: finding the element of the linear space of functions of single variables that captures most of the variance of U.

   This leads us to the idea of **Hájek projections**.

Projection Theorem

Consider a random variable T and a linear space $\mathcal{S}$ of random variables, with $\mathbf{E}S^2 < \infty$ for all $S \in \mathcal{S}$ and $\mathbf{E}T^2 < \infty$. [Write $T \in L_2(P)$, $\mathcal{S} \subset L_2(P)$, the Hilbert space of finite variance random variables defined on a probability space.] A **projection** $\hat{S}$ of T on $\mathcal{S}$ is a minimizer over $\mathcal{S}$ of $\mathbf{E}(T - S)^2$. [Picture]

Theorem: $\hat{S}$ is a **projection** of T on $\mathcal{S}$ iff $\hat{S} \in \mathcal{S}$ and, for all $S \in \mathcal{S}$, the error $T - \hat{S}$ is orthogonal to $\mathcal{S}$, that is,

$$\mathbf{E}(T - \hat{S})S = 0.$$

If $\hat{S}_1$ and $\hat{S}_2$ are projections of T onto $\mathcal{S}$, then $\hat{S}_1 = \hat{S}_2$ a.s.

Projection Theorem

Notice that if $\mathcal{S}$ contains constants, then $S = 1 \in \mathcal{S}$ shows that

$$\mathbf{E}(T - \hat{S}) = 0, \qquad \text{i.e., } \mathbf{E}T = \mathbf{E}\hat{S}.$$

Also, for all $S \in \mathcal{S}$, $S - \mathbf{E}S \in \mathcal{S}$, so

$$\text{Cov}(T - \hat{S}, S) = \mathbf{E}((T - \hat{S})(S - \mathbf{E}S)) = 0.$$

Projection Theorem Proof

Theorem: 1. $\hat{S} \in \mathcal{S}$ is a **projection** of T on $\mathcal{S}$ (minimizes $\mathbf{E}(T-S)^2$) iff, for all $S \in \mathcal{S}$, $\mathbf{E}(T-\hat{S})S = 0$.
2. If $\hat{S}_1$ and $\hat{S}_2$ are projections of T onto $\mathcal{S}$, then $\hat{S}_1 = \hat{S}_2$ a.s.

We can write the criterion, for any $S \in \mathcal{S}$ as

$$\begin{aligned}\mathbf{E}(T-S)^2 &= \mathbf{E}(T-\hat{S}+\hat{S}-S)^2 \\ &= \mathbf{E}(T-\hat{S})^2 + 2\mathbf{E}((T-\hat{S})(\hat{S}-S)) + (\hat{S}-S)^2.\end{aligned}$$

If $\mathbf{E}(T-\hat{S})S = 0$, then this is $\mathbf{E}(T-\hat{S})^2 + (\hat{S}-S)^2$, which is minimized for $S = \hat{S}$, and strictly minimized unless $\mathbf{E}(\hat{S}-S)^2 = 0$, so $\hat{S}$ is unique.

Projection Theorem Proof

If $\hat{S}$ is a projection, then

$$\mathbf{E}(T - \hat{S} - \alpha S)^2 = \mathbf{E}(T - \hat{S})^2 - 2\alpha \mathbf{E}(T - \hat{S})S + \alpha^2 \mathbf{E}S^2$$

is at least $\mathbf{E}(T - \hat{S})^2$ for any $S \in \mathcal{S}$ and any α. And this implies that $\mathbf{E}(T - \hat{S})S = 0$.

Projection Theorem

- Pythagoras theorem: $\mathbf{E}(T)^2 = \mathbf{E}(T - \hat{S} + \hat{S})^2 = \mathbf{E}(T - \hat{S})^2 + \mathbf{E}(\hat{S})^2$.
- If $\mathcal{S}$ contains constants, $\mathbf{E}(T) = \mathbf{E}(\hat{S})$ and $\mathrm{Var}(T) = \mathrm{Var}(T - \hat{S}) + \mathrm{Var}(\hat{S})$.
- So if $\mathcal{S}$ contains constants and $\hat{S}$ and T have the same variance, then $\hat{S} = T$ a.s.
- A similar property holds asymptotically...

Projections and Asymptotics

Consider $\mathcal{S}_n$ a sequence of linear spaces of random variables that contain the constants and that have finite second moments.

Theorem: For T_n with projections $\hat{S}_n$ on $\mathcal{S}_n$,

$$\frac{\mathrm{Var}(T_n)}{\mathrm{Var}(\hat{S}_n)} \to 1 \qquad \Longrightarrow \qquad \frac{T_n - \mathbf{E}T_n}{\sqrt{\mathrm{Var}(T_n)}} - \frac{\hat{S}_n - \mathbf{E}\hat{S}_n}{\sqrt{\mathrm{Var}(\hat{S}_n)}} \xrightarrow{P} 0.$$

Projections and Asymptotics: Proof

Define

$$Z_n = \frac{T_n - \mathbf{E}T_n}{\sqrt{\mathrm{Var}(T_n)}} - \frac{\hat{S}_n - \mathbf{E}\hat{S}_n}{\sqrt{\mathrm{Var}(\hat{S}_n)}}.$$

Clearly, $\mathbf{E}Z_n = 0$.

$$\begin{aligned}
\mathrm{Var}(Z_n) &= 2 - 2\frac{\mathrm{Cov}(T_n, \hat{S}_n)}{\sqrt{\mathrm{Var}(T_n)}\sqrt{\mathrm{Var}(\hat{S}_n)}} \\
&= 2 - 2\frac{\sqrt{\mathrm{Var}(\hat{S}_n)}}{\sqrt{\mathrm{Var}(T_n)}} \\
&\to 0,
\end{aligned}$$

where the second equality is because $\mathcal{S}$ contains constants, so $\mathrm{Cov}(T_n - \hat{S}_n, \hat{S}_n) = 0$, hence $\mathrm{Cov}(T_n, \hat{S}_n) = \mathrm{Var}(\hat{S}_n)$.