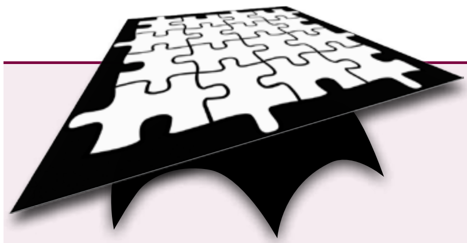

Student Puzzle Corner 16

Deadline October 23

It is the turn of a problem on statistics this time. This is the sixteenth problem in this problem series, and we must have had a good time, because we have already come to the end of three years since the series started. Here is the exact problem, the final one in 2016, and this one is going to be a pretty good teaser:

Let $X_1, X_2, \cdots, X_n \overset{\text{iid}}{\sim} F$, where $F(.)$ is a CDF on the real line; F is assumed to be unknown. Assume that F has a density $f(x)$, that $E_F(X^2) < \infty$, denote $E_F(X)$ by μ, and assume that $f(\mu) > 0$. Derive an asymptotically correct 95% confidence interval for $\theta = \theta(F) = F(\mu)$.

Note: You have to spend some time thinking if some additional control on the density f *is needed to make this go through.*

Student IMS members are invited to submit solutions (to bulletin@imstat.org with subject "Student Puzzle Corner"). The deadline is **October 23, 2016**. The names and affiliations of those submitting correct solutions, and the answer to the problem, will be published in the next issue. The Editor's decision is final.

Solution to Puzzle 15

Editor Anirban DasGupta writes:

Well done to Promit Ghosal at Columbia University [*right*], who sent a careful answer to this problem.

The problem asked was this. Take an integrable function f on the unit interval, and for x in the interval $[j/2^n, (j+1)/2^n)$, define $f_n(x)$ to be the average of f over $[j/2^n, (j+1)/2^n)$. Then, f_n converges pointwise to f for almost all x, and also converges to f almost uniformly.

First, the specific partition $[j/2^n, (j+1)/2^n), j = 0, 1, \cdots, 2^n - 1$ does not have much to do with the pointwise convergence; neither does the unit interval. We can conclude from real analysis that for any f which is merely locally integrable, for almost all x,

$$\lim_{h\to 0} \frac{1}{2h}\int_{x-h}^{x+h} |f(t) - f(x)|dt \to 0 \text{ as } h \to 0.$$

A point x satisfying this property is what analysts call a *Lebesgue point* of f. Almost all points x are Lebesgue points for a locally integrable function. The result generalizes to higher dimensions, and to well shaped shrinking neighborhoods.

Now, how does one prove this probabilistically? Consider the family of sets $\mathcal{A}_n = \{[j/2^n, (j+1)/2^n), j = 0, 1, \cdots, 2^n - 1\}$, and consider $\mathcal{F}_n$, the sigma-algebra generated by $\mathcal{A}_n$. Fix an x. Then in the probability space $([0, 1], \mathcal{B}, P)$, where P is Lebesgue measure on $[0, 1]$, the sequence $f_n(x)$ is a martingale for the filtration $\mathcal{F}_n$. This is a straight verification. It therefore follows from Doob's martingale convergence theorem that $f_n(x) \overset{\text{a.s.},P}{\to} f(x)$. Moreover, the convergence is almost uniform by *Egoroff's theorem*.

***Annals of Applied Statistics* 10(2):** continued from previous page

Clustering Chlorophyll-a satellite data using quantiles . . . . . CARLO GAETAN, PAOLO GIRARDI, ROBERTO PASTRES AND ANTOINE MANGIN 964

Asymmetric conditional correlations in stock returns . . . . . HUI JIANG, PATRICK W. SAART AND YINGCUN XIA 989

Regression analysis for microbiome compositional data . . . . . PIXU SHI, ANRU ZHANG AND HONGZHE LI 1019

Latent spatial models and sampling design for landscape genetics . . . . . EPHRAIM HANKS, MEVIN HOOTEN, STEVEN KNICK, SARA OYLER-MCCANCE, JENNIFER FIKE, TODD CROSS AND MICHAEL SCHWARTZ 1041

An imputation approach for handling mixed-mode surveys . . . . . SEUNGHWAN PARK, JAE KWANG KIM AND SANGUN PARK 1063

How strong is strong enough? Strengthening instruments through matching and weak instrument tests . . . . . LUKE KEELE AND JASON W. MORGAN 1086

Detecting abrupt changes in the spectra of high-energy astrophysical sources . . . . . RAYMOND K. W. WONG, VINAY L. KASHYAP, THOMAS C. M. LEE AND DAVID A. VAN DYK 1107

Correction

Bayesian structured additive distributional regression with an application to regional income inequality in Germany . . NADJA KLEIN, THOMAS KNEIB, STEFAN LANG AND ALEXANDER SOHN 1135