TENTAMEN Sannolikhetslära och inferensteori II, C1 (7.5 hp).
Måndagen den 27/10 2014, kl 8.00 - 13.00.
Ansvarig lärare: Måns Thulin, telefon 070-591 09 51. Kommer till salen kl 9.

Tillåtna hjälpmedel:
– Mendenhall, et. al.: Mathematical Statistics with Applications.
– Sydsaeter: Matematisk analys för ekonomer (eller motsvarande).
– Matematikhäfte från B-nivån (Lars Forsberg)
– Miniräknare.

1. (30 p) De kontinuerliga slumpvariablerna X och Y har täthetsfunktionen

$$f(x,y) = 1/3, \qquad x \leq y \leq x+1, \quad 0 \leq x \leq 3.$$

   (a) (5 p) Är X och Y oberoende? Motivera.
   (b) (10 p) Beräkna $E(Y|X=x)$.
   (c) (15 p) Beräkna kovariansen mellan X och Y.

2. (25 p) Den kontinuerliga slumpvariabeln Y har täthetsfunktionen:

$$f(y) = \frac{1}{2}(1+\theta y), \qquad -1 < y < 1, \quad -1 < \theta < 1.$$

   (a) (15 p) Härled en väntevärdesriktig skattare av θ baserad på n oberoende observationer av Y.
   (b) (10 p) Är skattaren du får fram i (a) konsistent? Motivera svaret.

3. (15 p) Låt $X_1, X_2, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_n$ vara oberoende $Bin(1,p)$-fördelade slumpvariabler, med $1 \leq n_1 < n/2$ och $n_2 = n - n_1$. Då gäller att $X = \sum_{i=1}^n \sim Bin(n,p)$. I en nyligen publicerad artikel[1] föreslås att man kan få bättre konfidensintervall för parametern p i binomialfördelningen genom att i formeln för konfidensintervallet ersätta observationen X med $\tilde{X}$, definerad som

$$\tilde{X} = \frac{n}{2}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1} X_i + \frac{1}{n_2}\sum_{i=n_1+1}^{n} X_i\Big).$$

[1] Decrouez, G., Hall, P. (2014). Split sample methods for constructing confidence intervals for binomial and Poisson parameters. *Journal of the Royal Statistical Society: Series B*, **76**, 949–975.

(a) (5 p) Bestäm den momentgenererande funktionen för $\tilde{X}$.

(b) (10 p) I en annan artikel[2] visades att man kan tolka det som att $\tilde{X}$ är lika med X plus ytterligare en slumpvariabel. Låt $Z \sim HypGeo(n, X, n_1)$. Visa att $\tilde{X}$ har samma fördelning som $X + Y$ där

$$Y = \frac{n}{2n_1}Z - \frac{n}{2n_2}Z + \frac{n_1 - n_2}{2n_2}X.$$

Ledning: börja med att undersöka kopplingen mellan Z och de två summorna i definitionen av $\tilde{X}$.

4. (30 p) Låt $Y_1, Y_2, \ldots, Y_n$ vara ett slumpmässigt stickprov från en fördelning med täthetsfunktion

$$f(y) = (1+\theta)\, y^{\theta},\ \theta > -1,\ 0 < y \leq 1$$

(a) (15 p) Antag att man vill utföra hypotestestet

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_a &: \theta = \theta_a \end{aligned}$$

där $\theta_a > \theta_0$. Bestäm det bästa kritiska området, dvs det som har den största styrkan i punkten $\theta = \theta_a$. Visa att det test som använder detta kritiska område är ett UMP-test för mothypoteser av typen $H_a : \theta > \theta_0$. *Ledning*: Logaritmering är ej nödvändig.

(b) (15 p) Antag att man vill testa

$$\begin{aligned} H_0 &: \theta = 0 \\ H_a &: \theta > 0 \end{aligned}$$

med en signifikansnivå på $\alpha = 0.05$ (dvs på 5%-nivån). Till vår hjälp har vi endast en observation av Y. Ange det kritiska området (förkastelseområdet) och testets styrkefunktion.

[2] Thulin, M. (2014). On split sample and randomized confidence intervals for binomial proportions. *Statistics & Probability Letters*, **92**, 65–71.