Tentamen Inleiding Kansrekening en Statistiek
2 februari 2017, 13.30-16.30

- Laat duidelijk zien hoe je aan je antwoorden bent gekomen.
- Maak opgaven 1 en 2 samen op een apart blad. Maak opgaven 3, 4, 5 allen op een apart blad.

---

1. (10 pt) *Hypothese:* Als A en B onafhankelijke kansgebeurtenissen zijn, dan zijn A^C en B^C ook onafhankelijke kansgebeurtenissen.

   Geef een bewijs van bovenstaande Hypothese, of geef een tegenvoorbeeld.

2. (15 pt) Neem aan dat X en Y onafhankelijke kansvariabelen zijn, die allebei $Exp(\lambda)$ verdeeld zijn. Definieer $M = max\{X, Y\}$ en $Z = X + (Y/2)$.

   (a) (7 pt) Bepaal de kansdichtheid van M.

   (b) (8 pt) Bepaal de kansdichtheid van Z.

3. (25 pt) Het is 1813. Op Pemberley House wordt een groot, formeel bal gehouden. Er zijn k heren en n dames. Elke heer kiest een dame, willekeurig en uniform. Anders gezegd, de kans dat heer i dame j kiest is $1/n$, voor alle $i = 1, 2, \ldots, k$ en $j = 1, 2, \ldots, n$. Als meerdere heren zich bij een dame melden, kiest zij haar favoriet en gaan de afgewezen heren weer naar de kant. NB: er zullen waarschijnlijk dames zijn die door niemand worden gekozen. Ook zij gaan naar de kant.

   (a) (5 pt) Zij D_i het aantal heren dat zich bij dame i meldt. Beargumenteer dat $D_i \sim Bin(k, 1/n)$.

   (b) (10 pt) Neem aan dat $k = \alpha n$, voor een $\alpha > 0$. Laat zien dat $D_i \sim Pois(\alpha)$, in de limiet dat $n \to \infty$.

   (c) (10 pt) Zij F_n de fractie dansende dames (aantal gekozen dames gedeeld door n). Neem aan dat n groot genoeg is om de benadering van (b) te gebruiken. Wat is de verwachting van F_n?

4. (30 pt) De kansvariabelen X_i, $i = 1, 2, \ldots, n$ zijn onderling onafhankelijk en allen uniform verdeeld op $[0, \theta]$ met $\theta > 0$. Zij $\overline{X}_n = (X_1 + X_2 + \ldots X_n)/n$.

   (a) (5 pt) Construeer een zuivere schatter T voor θ, op basis van $\overline{X}_n$.

   (b) (10 pt) Het is bekend dat de Maximum Likelihood Estimator voor θ gegeven wordt door $M = max\{X_1, \ldots, X_n\}$. Bereken $MSE[T]$ en $MSE[M]$.

   (c) (5 pt) Voor grote n heeft $Z = \sqrt{n}(T - \mu)/\sigma$ bij benadering een $N(0, 1)$ verdeling. Hierbij is $\mu = E[T]$ en $\sigma^2 = Var[T]$. Geef de waarden van μ en σ in termen van θ.

   (d) (10 pt) Vind, door gebruik te maken van de normale benadering uit (c), een 95% betrouwbaarheidsinterval voor θ. NB: $P(-1,96 \leq Z \leq 1,96) = 0,95$, als Z standaard normaal verdeeld is.

5. (20 pt) Zij X een discrete en eindige kansvariabele met uitkomstenruimte $\{x_1, \ldots, x_n\}$, met bijbehorende kansen $P(X = x_k) = p_k > 0$, $k = 1, \ldots, n$.
   De *entropie* van X wordt gedefinieerd door $H(X) = \sum_{k=1}^{n} p_k log(1/p_k)$.
   Als $Z = \{z_1, \ldots, z_n\}$, dan schrijven we $E_X[Z] = \sum_{k=1}^{n} p_k z_k$.
   Je mag in deze opgave de volgende vorm van de ongelijkheid van Jensen gebruiken: als $g(x)$ een concave functie is, dan is $E_X[g(Z)] \leq g(E_X[Z])$. Hierbij is $g(Z) = \{g(z_1), \ldots, g(z_n)\}$.

   (a) (5 pt) Laat zien dat $H(X) \leq log(n)$. Hint: gebruik Jensen's ongelijkheid voor $Z = \{1/p_1, \ldots, 1/p_n\}$. Bedenk zelf een geschikte functie $g(x)$.

   (b) (10 pt) Neem aan dat X en Y onafhankelijke, discrete en eindige kansvariabelen zijn met identieke kansverdeling. Bewijs dat $P(X = Y) \geq e^{-H(X)}$. Hint: beschouw een kansvariabele W met uitkomsten $\{p_1, \ldots, p_n\}$ en kansen $P(W = p_k) = p_k$, $k = 1, \ldots, n$.

   (c) (5 pt) Combineer (a) en (b) tot een ondergrens voor $P(X = Y)$ die alleen van n afhangt.