**Johdatus todennäköisyyslaskentaan, kevät 2014**
**Helsingin yliopisto, Matematiikan ja tilastotieteen laitos**
**Harjoitus 6 – Palautuspäivä 19.2.2014**

Palauta ratkaisusi 3. kerroksen C-käytävän muovilaatikkoon. Käytä kurssitunnusta, älä opiskelijanumeroa!

## Tehtäväsarja I: Ehdolliset todennäköisyydet

**1.** Olkoot $A$ ja $B$ sellaisia tapahtumia, että $0 < P(B) < 1$, ts. että sekä $B$ että sen komplementti ovat mahdollisia. **Todista täsmällisesti** seuraava väite:

Jos luvuista $P(A)$, $P(A \mid B)$ ja $P(A \mid B^c)$ mitkä tahansa kaksi ovat yhtäsuuret, niin kolmaskin on yhtäsuuri niiden kanssa.

Vihje: Kokonaistodennäköisyyden kaava. Oleta, että jotkut kaksi mainituista kolmesta luvusta ovat samat (tähän on kolme erilaista vaihtoehtoa) ja laske sen perusteella kolmas niistä.

**2.** Jatkoa edelliseen tehtävään. Äsken tuli itse asiassa todistettua, että nuo kolme todennäköisyyttä ovat joko *kaikki* erisuuria (= kolme eri lukua) tai *kaikki* yhtäsuuria. Keksi jokin konkreettinen tapaus (esim. nopanheitosta, korttipakasta, populaatiosta tms.) jossa ne ovat kaikki erisuuria ja laske ne.

## Tehtäväsarja II: Tiheysfunktio, kertymäfunktio ja odotusarvo

***3.** Kahden riippumattoman Tas$(0, 1)$-satunnaismuuttujan (esim. $X$ ja $Y$) summa on $Z$. Tiedetään (vrt. Tuominen s. 72), että $Z$:llä on tiheysfunktio

$$f(z) = \begin{array}{ll} z, & \text{kun } 0 < z \leq 1, \\ 2 - z, & \text{kun } 1 < z < 2, \\ 0 & \text{muualla.} \end{array}$$

Laske $E(Z)$ ja $P(0.5 < Z < 1.5)$.

Vihje: Odotusarvon voi laskea kahdella tavalla: lineaarisuuden (lause 3.1.1) perusteella tai tiheysfunktiosta. Tuloksen pitäisi olla sama!

**4.** Satunnaismuuttujalla $X$ on ns. kaksipuolinen eksponenttijakauma eli Laplacen jakauma: sen tiheysfunktio on

$$f(x) = \frac{1}{2} \exp(-|X|).$$

(a) Tarkista integroimalla, että kyseessä on todella tiheysfunktio.

(b) Laske $E(X)$.

(c) Laske $P(|X| > 1)$.

**5.** Satunnaismuuttujalla $\theta$ on tasajakauma välillä $(0, 2\pi)$. Tutkitaan satunnaismuuttjien $X = \cos\theta$ ja $Y = \sin\theta$ jakaumaa empiirisesti.

(a) Muodosta sadan kappaleen satunnaisotos $\theta$:n jakaumasta, laske vastaavat $X$:n ja $Y$:n arvot, ja piirrä näin saadut pisteet esim. tämäntapaisella Matlab-komennolla: `plot(X,Y,'.')`. Mihin tason osaan pisteet $(X, Y)$ sijoittuvat?

Vihje: Matlabissa kosini- ja sinifunktiota voi soveltaa kokonaiseen vektoriin, tulos on vastaavasti vektori.

(b) Muodosta tuhannen kappaleen satunnaisotos $\theta$:n jakaumasta ja piirrä histogrammi $X$:n arvoista. Vaikuttaako jakauma tasaiselta? Mitä arvelet syyksi?

**6.** Jatkoa edelliseen tehtävään.

(a) Olkoon $x$ jokin luku välillä $(-1, 1)$. Millainen on $\theta$:n oltava, jotta $X \leq x$? (Vihje: Tutki esim. kosinifunktion kuvaajaa.)

(b) Mikä on a-kohdassa mainitun tapahtuman (epäyhtälön toteutumisen) todennäköisyys?

(c) Määrittele edellisten kohtien perusteella $X$:n kertymäfunktio ja tiheysfunktio. Piirrä tiheysfunktio (koneella tai käsin).

Vihje: arkuskosinin derivaatta on

$$\frac{d}{dx} \arccos x = -\left(\frac{1}{\sqrt{1 - x^2}}\right).$$

***7.** Jatkoa edelliseen tehtävään. $X$:n jakauman täsmällisen muodon selvittäminen onnistui mutta oli hiukan työlästä. Jos olisimme olleet kiinnostuneita ainoastaan $X$:n odotusarvosta, olisimme voineet käyttää lausetta 3.1.8 (sivu 82), jolla saadaan muunnoksen $X = \cos\theta$ odotusarvo, kun alkuperäisen satunnaismuuttujan $\theta$ tiheysfunktio tunnetaan. Laske tällä tavalla $E(X)$.

### Tehtäväsarja III: Multinomikerroin ja multinomijakauma

**8.** Olkoot $X_1, \ldots, X_{99}$ riippumattomia ja $X_i \sim \mathrm{Tas}(0, 1)$ (kaikilla $i = 1, \ldots, 99$). Luvuista piirretään hyvin karkea histogrammi, jossa on vain kolme pylvästä:

- Pylvään 1 korkeus, merkitään $N_1$, on välille $(0, 1/3]$ osuneiden lukujen lukumäärä.
- Pylvään 2 korkeus, merkitään $N_2$, on välille $(1/3, 2/3]$ osuneiden lukujen lukumäärä.
- Pylvään 3 korkeus, merkitään $N_3$, on välille $(2/3, 1)$ osuneiden lukujen lukumäärä.

Mikä on luvun $N_1$ jakauma ja odotusarvo? Entä lukujen $N_2$ ja $N_3$? Mikä on todennäköisyys, että $N_1$ osuu täsmälleen odotusarvoonsa?

**9.** Jatkoa edelliseen. Mikä on todennäköisyys, että histogrammi on täysin tasainen, ts. kolme pylvästä ovat yhtä korkeat eli $N_1 = N_2 = N_3$? Vihje: Pylväiden korkeuksilla on multinomijakauma (ks. kurssisivulta tarkempaa selitystä.) Voit laskea multinomikertoimen tarkasti (kertomien avulla) tai esim. Stirlingin approksimaatiolla.

**10.** Jatkoa edelliseen. Ovatko satunnaismuuttujat $N_1$, $N_2$ ja $N_3$ riippumattomat?

**Tehtäväsarja IV: Varianssi ja suurten lukujen laki**
Tutustu Tuomisen lukuihin 3.2 ja 3.5.

***11.** (vrt. Tuominen 3:10) Suorakulmion sivujen todelliset pituudet ovat 10 ja 20 senttimetriä. Mittauksessa tapahtuu virhe siten, että mittaustulokset ovat $A = a + X$ ja $B = b + Y$, missä virheet $X$ ja $Y$ ovat riippumattomat ja tasajakautuneet välillä $(-1, 1)$.

Olkoon $Z = AB - ab$ eli suorakulmion mittausten mukaisen pinta-alan ja oikean pinta-alan erotus, ts. mitatun pinta-alan virhe. Laske $Z$:n odotusarvo ja varianssi.

**12.** Painotettua kolikkoa, jossa kruunan todennäköisyys on $p = 1/3$, heitetään $n$ kertaa. Olkoon $Y$ kruunien lukumäärä ja $X = Y/n$ kruunien suhteellinen frekvenssi. Laske $X$:n odotusarvo, varianssi ja keskihajonta.

**13.** Jatkoa edelliseen. Olkoon $\epsilon = 0.01$. Toivomme, että $Y$ olisi $\epsilon$:n tarkkuudella likiarvo oikealle todennäköisyydelle, ts. toivomme, että toteutuu epäyhtälö $|X - p| < \epsilon$. Sen komplementtitapahtumaa $|X - p| \geq \epsilon$ kutsumme *liian suureksi virheeksi*, ja merkitään kirjaimella $L$. Laske Tsebysevin epäyhtälön (lause 3.2.8 sivu 85) perusteella yläraja todennäköisyydelle $P(L)$.

**14.** Jatkoa edelliseen. Kuinka suuri on $n$:n oltava, jotta Tsebysevin epäyhtälön nojalla $P(L) < 0.05$?

**15.** Jatkoa edelliseen. Simuloi painotetun kolikon heittoa $n$ kertaa ja tutki, kuinka suuri oli toteutunut virhe $|X - p|$. Oliko virhe asettamamme tarkkuusvaatimuksen mukainen eli pienempi kuin $\epsilon$?

**16.** (Monte Carlo -integrointi) Kurssisivulta löytyy MATLAB-apufunktio `kuvio`, joka tutkii annetusta tason pisteestä, onko se erään kuvion sisäpuolella vai ei. Toisin sanoen kutsu `kuvio(x,y)` palauttaa arvon 1, jos piste on kuvion sisäpuolella, ja 0, jos ulkopuolella tai reunalla. ($x$ ja $y$ voivat olla myös vektoreita, jolloin funktio tutkii kutakin pistettä erikseen.)

Tiedämme, että kuvio sisältyy neliöön $(-1, 1)^2$, toisin sanoen sen kaikkien sisäpisteiden sekä x- ja y-koordinaatit ovat välillä $(-1, 1)$. Tiedämme, että kyseisen neliön pinta-ala on $a = 4$.

Yritetään approksimoida kuvion pinta-alaa seuraavasti:

- Arvotaan $n = 1000\,000$ em. neliön pistettä $(X, Y)$, missä $X$ ja $Y$ ovat riippumattomasti tasajakautuneet välillä $(-1, 1)$.
- Jokaisesta pisteestä tutkitaan funktiolla `kuvio`, onko piste kuvion sisällä.
- Olkoon $k$ kuvioon osuneiden pisteiden lukumäärä. Arvioidaan, että kuvion pinta-ala on $\approx a \cdot (k/n)$.

Suorita approksimaatio. Millaisen pinta-ala-arvion sait?

**Tehtäväsarja V: Normaalijakauma ja normaaliapproksimaatio**
Tutustu Tuomisen lukuun 4.

***17.** (vrt. Tuominen 2:51) Henkilöt A ja B ovat päättäneet tavata rautatieasemalla. He lähtevät samanaikaisesti bussilla päätepysäkeiltä, A Kontulasta, B Lauttasaaresta. Bussien kulkuajat minuutteina ovat riippumattomat ja jakaumiltaan: Kontulan bussi $N(25, 8^2)$ ja Lauttasaaren bussi $N(30, 6^2)$.

Merkitään $X$:llä satunnaismuuttujaa, joka ilmaisee, montako minuuttia A:n jälkeen B saapuu; negatiivinen arvo merkitsee, että B saapuukin ensin.

(a) Mikä on $X$:n jakauma?

(b) Mikä on tn, että B saapuu ensin?

Vihje: Normaalijakauman kertymäfunktio löytyy taulukoista (esim. Tuomisen kirjan takana), monista laskimista, ja Matlabista funktiolla `normcdf`.

**18.** Jatkoa edelliseen tehtävään.

(a) Mikä on tn, että A joutuu odottamaan yli 5 minuuttia?

(b) Mikä on tn, että B joutuu odottamaan yli 5 minuuttia?

(c) Mikä on tn, että jompikumpi joutuu odottamaan yli 5 minuuttia?

**19.** Koetta, jossa heitetään kuutta arpakuutiota, toistetaan 3000 kertaa. Tarkastellaan tapahtuman $A$ = “kaikki kuusi noppaa osoittavat eri pistelukua” esiintymistä. Laske normaaliapproksimaatiolla tn, että $A$ esiintyy toistokokeessa yli 50 kertaa.