Tõenäosusteooria ja statistika II

# EKSAM

5. jaanuaril 2022

NIMI: ……………………………………………………………

**Ülesanne 1**

(9 punkti)

Kahe juhusliku suuruse, *X* ja *Y*, ühisjaotust kirjeldab järgmine tabel:

| $X \setminus Y$ | 1 | 2 | 3 |
|---|---|---|---|
| 10 | 0,1$p$ | 0,2$p$ | 0,3$p$ |
| 20 | 0,1$p$ | 0,05$p$ | 0,05$p$ |
| 30 | 1-$p$ | 0,1$p$ | 0,1$p$ |

Kus parameeter $0 \le p \le 1$. Jaotuse parameetri väljaselgitamiseks tehti ka kolm mõõtmist. Saadud mõõtmistulemused olid järgmised:

| Mõõtmise nr | *X* | *Y* |
|---|---|---|
| 1 | 10 | 1 |
| 2 | 10 | 3 |
| 3 | 30 | 1 |

a. (3 punkti) Leia juhusliku suuruse *Y* tinglik jaotus juhul kui *X*=20 (jaotus võib sisaldada tundmatut parameetrit *p* – parameetri väärtust ei pea asendama hinnanguga):

$Y|X$=20

| $y_i$ | 1 | 2 | 3 |
|---|---|---|---|
| P($Y=y_i$) | 0,5 | 0,25 | 0,25 |

b. (3 punkti) Leia suurima tõepära hinnang parameetrile *p* (pane kirja ka hinnangu tuletuskäik: tõepärafunktsioon jne).

$L = 0{,}1p \cdot 0{,}3p \cdot (1-p)$

$l = \log(0{,}1) + \log(p) + \log(0{,}3) + \log(p) + \log(1-p)$

$l' = 1/p + 1/p - 1/(1-p)$

$l' = 0 \quad \Rightarrow \quad 2/p = 1/(1-p) \;|\cdot p(1-p)$

$2(1-p) = p$

$2 = 3p \quad \Rightarrow \hat{p}=2/3$

c. (3 punkti) Millise hinnangu saaksime parameetrile p siis, kui X-tunnuse väärtuseid poleks mõõdetud (oleksime näinud vaid Y-tunnuse väärtuseid: 1; 3; 1). Pane kirja ka hinnangu tuletuskäik!

*Y*-i marginaaljaotus:

| $y_i$ | 1 | 2 | 3 |
|---|---|---|---|
| P($Y=y_i$) | 1-0,8$p$ | 0,35$p$ | 0,45$p$ |

$L = (1-0{,}8p)^2 \cdot 0{,}45p$

$l = 2\log(1-0{,}8p) + \log(0{,}45) + \log(p)$ Maksimumi leidmiseks $l'$=0:

$2/(1-0{,}8p) \cdot (-0{,}8) + 1/p = 0 \quad |\cdot p(1-0{,}8p)$

$-1{,}6p + (1-0{,}8p) = 0$

$-2{,}4p + 1 = 0 \quad \Rightarrow \hat{p}=1/2{,}4 = 0{,}41666...$

**Ülesanne 2**
**(9 punkti)**

Juhuslike suuruste *X* ja *Y* ühisjaotust kirjeldab järgmine tihedusfunktsioon:

$$f_{X,Y}(x,y) = \begin{cases} -x + 2y - 2xy + 1, & \text{kui } 0 \le x \le 1,\ 0 \le y \le 1, \\ 0 & \text{mujal.} \end{cases}$$

Leia P(*X*<*Y*), E(*XY*), E(*X*)!

$P(X < Y) = \int_0^1 \int_0^y (-x + 2y - 2xy + 1)\, dx\, dy$

$$= \int_0^1 (-x^2/2 + 2yx - x^2y + x)\,|_0^y dy$$

$$= \int_0^1 \left(-\frac{y^2}{2} + 2y^2 - y^3 + y\right) dy$$

$$= \int_0^1 \left(-\frac{y^3}{6} + \frac{2}{3}y^3 - \frac{y^4}{4} + \frac{y^2}{2}\right) dy$$

$$= -\frac{1}{6} + \frac{2}{3} - \frac{1}{4} + \frac{1}{2} = \frac{-2+8-3+6}{12} = \frac{9}{12} = \frac{3}{4} = 0{,}75$$

E(*XY*) $= \int_0^1 \int_0^1 xy(-x + 2y - 2xy + 1)\, dx\, dy$

$$=\int_0^1 -\frac{x^3y}{3} + x^2y^2 - \frac{2x^3y^2}{3} + \frac{x^2y}{2}\ |_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(-\frac{y}{3} + y^2 - \frac{2y^2}{3} + \frac{y}{2}\right) dy$$

$$= \left(-\frac{y^2}{6} + \frac{y^3}{3} - \frac{2y^3}{9} + \frac{y^2}{4}\right)|_{y=0}^{y=1} = \left(-\frac{1}{6} + \frac{1}{3} - \frac{2}{9} + \frac{1}{4}\right) = \frac{-6+12-8+9}{36}$$=7/36 = 0,19444…

$f_X(x) = \int_0^1 (-x + 2y - 2xy + 1)\, dy = (-xy + y^2 - xy^2 + y)|_{y=0}^{y=1} = -x + 1 - x + 1$

$$= 2 - 2x, \text{kui } 0 \le x \le 1$$

$E(X) = \int_0^1 x(2 - 2x)dx = x^2 - \frac{2}{3}x^3|_{x=0}^{x=1}$=1-2/3 = 1/3

**Ülesanne 3**

(8 punkti)

Eesti elanikest 26% on pensionärid. Soovitakse teada, kas riigi poole infopäringutega pöördujate seas on pensionäride osakaal samasugune kui elanikkonnas tervikuna. Küsitleti 10 (juhuslikult valitud) infopäringu tegijat ja selgitati välja kas nad on pensionärid või mitte (selgus, et 1 neist oli pensionär). Testi järgmiseid hüpoteese:

$H_0$: Infopäringu teinute seas on pensionäride osakaal sama mis rahvastikus tervikuna, P(pensionär)=0,26

$H_1$: P(pensionär) ≠ 0,26.

Kasuta teststatistikuna pensionäride arvu valimis (ja olulisuse nivood 0,05). Vaata ka alljärgnevaid tabeleid:

X~B(n=10; p=0,26)

| x | P(X=x) |
|---|---|
| 0 | 0.049240 |
| 1 | 0.173005 |
| 2 | 0.273535 |
| 3 | 0.256285 |
| 4 | 0.157581 |
| 5 | 0.066439 |
| 6 | 0.019453 |
| 7 | 0.003906 |
| 8 | 0.000515 |
| 9 | 0.000040 |
| 10 | 0.000001 |

X~B(n=10; p=0,1)

| x | P(X=x) |
|---|---|
| 0 | 0.3486784401 |
| 1 | 0.3874204890 |
| 2 | 0.1937102445 |
| 3 | 0.0573956280 |
| 4 | 0.0111602610 |
| 5 | 0.0014880348 |
| 6 | 0.0001377810 |
| 7 | 0.0000087480 |
| 8 | 0.0000003645 |
| 9 | 0.0000000090 |
| 10 | 0.0000000001 |

Leia testi p-väärtus (3 punkti): … 1-0,273535-0,256285 = 0,47 ……

Milline on testi otsus, kas $H_0$ või $H_1$ (1 punkt): ……… $H_0$ …………

Milline oleks testi võimsus kui 10% infopärijatest oleksid pensionärid (3 punkt)?

**…** 0,0001377 + 0,0000008748 + 0,0000003645 + 0,000000009 + 0,0000000001 = 0.000139 **…**

Kas oled tõenäolisemalt teinud I-liiki või II-liiki vea (1 punkti):

…………………… II-liiki vea (I-liiki viga ei saa teha kui jääme $H_0$ juurde) …………………………

**Ülesanne 4**

(8 punkti)

Tootja valmistab tooteid kolmel meetodil: meetod A, meetod B ja meetod C. Varasemast on teada, et meetodil A valmistatud tooted lagunevad keskmiselt poole kiiremini kui meetodil B valmistatud tooted ja meetodil C valmistatud tooted peavad keskmiselt vastu 2 korda kauem kui meetodil B valmistatud tooted. Võib eeldada, et kõigil kolmel meetodil valmistatud toodete eluead on eksponentjaotusega.

Nelja toote jaoks mõõdeti ka nende vastupidavust (kui kaua nad töötavad). Saadi järgmised tulemused:

| Valmistusmeetod | Eluiga (kaua oli töökorras, kuudes) |
|---|---|
| A | 12 |
| A | 16 |
| A | 21 |
| C | 34 |

Leia suurima tõepära hinnang meetodil B valmistatud toodete keskmisele elueale!

Märkus: Eksponentjaotuse Exp(λ) tihedusfunktsioon on kujul $f(x) = \lambda e^{-\lambda x}$, kui $x \geq 0$ (muidu on tihedusfunktsiooni väärtuseks 0).

Meetod A: $f(x) = 2\lambda e^{-2\lambda x}$

Meetod B: $f(x) = \lambda e^{-\lambda x}$

Meetod C: $f(x) = ½\, \lambda e^{-1/2\, \lambda x}$

Valimi ühine tõepära:

$$L = 2\lambda e^{-2\lambda\, 12}\, 2\lambda e^{-2\lambda\, 16}\, 2\lambda e^{-2\lambda\, 21}\ ½\, \lambda e^{-\lambda\, 34/2}$$

$$L = 4\, \lambda^4 \exp(-2\lambda(12+16+21+34/4))$$

$$l = \log(4)+4\log(\lambda) -2\lambda(12+16+21+34/4)$$

Leiame maksimumi: l’=0 =>

$$4/\, \lambda = 2(12+16+21+34/4)$$

$$\hat{\lambda} = 2/(12+16+21+34/4) = 0{,}03478\ldots$$

Meetodil B valmistatud toote keskmine eluiga on seega (hinnatud suurima tõepära meetodi abil):

$$\widehat{E(X)} = 1/\, \hat{\lambda} = 1/\, 0{,}03478.. = 28{,}75$$

## Ülesanne 5

(4 punkti)

Otsitakse imetoitu, mida enne eksamit manustades tõuseks eksami keskmine hinne. Katsetati 8-t erinevat sööki ja võrreldi neid tudengeid, kes enne eksamit antud toitu sõid nendega kes ei söönud enne eksamit midagi (*t*-test sõltumatute valimite jaoks). Saadud p-väärtused (olulisustõenäosused) olid järgmised:

| Toit | p-väärtus |
|---|---|
| Kaerahelbepuder: | 0.0031 |
| *Bœuf Bourgignon*: | 0.0154 |
| Läätseleem: | 0.0118 |
| Kiirnuudlid: | 0.0121 |
| Must leib: | 0.0070 |
| Soolaheeringas: | 0.0353 |
| Kisla: | 0.0079 |
| Praemuna peekoniga: | 0.4849 |

Milliste toitude eksamisooritust abistavast võimest võime rääkida, kui tahame, et meie poolt mainitud toitude seas poleks ekslikult nimekirja valitud toite suurema tõenäosusega kui 0,05?

Selgita ka, kuidas tegid oma valikud!

Kuna teame kõigi testide p-väärtuseid siis on mõistlik kasutada Bonferron-Holmi meetodit (mis on võimsam kui Bonferrroni meetod).

Esmalt järjestame p-väärtused:

0,0031 < 0,007 < 0,0079 < 0,0118 < 0,0121 < 0,0154 < 0,0353 < 0,4849

Kontrollime:

0,0031 < 0,05/8 = 0,00625 => $H_1$

0,007 < 0,05/7 = 0,00714 => $H_1$

0,0079 < 0,05/6 = 0,00833 => $H_1$

0,0118 > 0,05/5 = 0,001 => $H_0$ (ja ka kõigi järgnevate testide puhul jääme $H_0$ juurde).

Seega kaerahelbeputru, leiba või kislat hommikul süües saab tõestatavalt mõjutada eksami hinnet.

## Ülesanne 6

(4 punkti)

Negatiivse binoomjaotusega juhusliku suuruse $X$~NB($r,p$) momente genereeriv funktsioon on kujul

$$M_X(t) = \left(\frac{1-p}{1-pe^t}\right)^r, \text{kui } t < -\log(p).$$

Tõesta, et kahe sõltumatu negatiivse binoomjaotusega juhusliku suuruse $X_1$~NB($r_1$; $p$) ja $X_2$~NB($r_2$; $p$) summa on ka negatiivse binoomjaotusega, $X_1+X_2$~NB($a,b$). Kuidas avalduvad jaotuse parameetrid $a$ ja $b$ esialgsete juhuslike suuruste jaotusparameetrite $r_1$, $r_2$ ja $p$ kaudu?

Sõltumatute juhuslike suuruste summa momente genereeriv funktsioon on liidetavate momente genereerivate funktsioonide korrutis:

$$M_{X_1+X_2}(t) = \left(\frac{1-p}{1-pe^t}\right)^{r_1}\left(\frac{1-p}{1-pe^t}\right)^{r_2} = \left(\frac{1-p}{1-pe^t}\right)^{r_1+r_2}$$

Ehk summa $X_1+X_2$ jaotuseks on negatiivne binoomjaotus, $X_1+X_2$~NB($r_1+r_2$, $p$)

**Ülesanne 7**

(4 punkti)

Ühes piirkonnas leidub kolme metsiku kalkuni alamliiki (50% antud piirkonna metsikutest kalkunitest kuuluvad alamliiki Gould, 20% kuuluvad Rio Grande alamliiki ja 30% kuuluvad Florida alamliiki). Kõigi alamliikide kohta on teada kaalu keskväärtus ja dispersioon:

| Alamliik | kaalu keskväärtus | kaalu dispersioon |
|---|---|---|
| Gould | 10 kg | 2 |
| Rio Grande | 9 kg | 2 |
| Florida | 8 kg | 0,5 |

Leia antud piirkonnas elavate kalkunite keskmine kaal ja kaalu dispersioon! Selgita, kuidas teostasid arvutused (milliseid valemeid kasutades).

$E(X) = E\,E(X|Y) = 10\cdot 0{,}5 + 9\cdot 0{,}2 + 8\cdot 0{,}3 = 5+1{,}8+2{,}4 = 9{,}2$ kg

$D(X) = E\,D(X|Y) + D\,E(X|Y)$

$E\,D(X|Y) = 2\cdot 0{,}5+2\cdot 0{,}2+0{,}5\cdot 0{,}3 = 1{,}55$

$D\,E(X|Y) = (10-9{,}2)^2 \cdot 0{,}5 + (9-9{,}2)^2 \cdot 0{,}2 + (8-9{,}2)^2 \cdot 0{,}3 = 0{,}76$

$D(X) = 1{,}55 + 0{,}76 = 2{,}31$

**Ülesanne 8**

(4 punkti)

Juhuslik suurus $X$ on t-jaotusega, $X \sim t(df)$.

a) Mis jaotusega on juhuslik suurus $X^2$?

b) Mis jaotusega on juhuslik suurus $1/X^2$?

Vastuses ootan nii jaotuste perekonna nime kui ka parameetri või parameetrite väärtuseid, samuti tahan näha põhjendust – miks transformeeritud juhusliku suuruse jaotus just selline on!

Kui $Y \sim N(0,1)$ ja $Z \sim X^2(df)$ on sõltumatud juhuslikud suurused, siis $X = \frac{Y}{\sqrt{Z/df}} \sim t(df)$ ja

$$X^2 = \frac{Y^2/1}{Z/df}.$$

Kuna $Y^2$ on hii-ruut jaotusega juhuslik suurus (vabadusastmete arvuga 1) siis $X^2$ on F-jaotusega juhuslik suurus parameetritega $df_1=1$, $df_2=df$ (kui jagada kahte sõltumatut hii-ruut jaotusega juhuslikku suurust nende vabadusastmete arvudega ja saadud jagatised jagada, siis on tulemuseks F-jaotusega juhuslik suurus).

F-jaotuse omadustest (loeng 8, slaid 30) tulenevalt – F-jaotusega juhusliku suuruse pöördväärtus on ka F-jaotusega – saame:

$$1/X^2 \sim F(df, 1).$$