HerTentamen Inleiding Kansrekening en Statistiek (WISB161) 2017-2018

19 april 2018

Opgave 1 Stel dat er verkiezingen zijn geweest waarbij N individuen gestemd hebben. Er zijn 2 kandidaten, kandidaat A en kandidaat B. Kandidaat A heeft 53% van de stemmen behaald en kandidaat B 47%. Voor de verkiezingen is er een representatieve steekproeven gehouden van grootte $n = 1000$ en veronderstel dat iedereen in de peiling correct aangeeft op welke kandidaat hij/zij zal stemmen. Definieer de stochastische variabelen p_A en p_B als de fractie van de deelnemers in de steekproef die, respectievelijk, op kandidaat A en kandidaat B zullen stemmen. Ga ervan uit dat $N \gg n$.

a 8pt) Wat is de variantie van p_A?

ANTWOORD: Definieer de Bernoulli-stochast X_i door
$X_i = \begin{cases} 1 \text{ als de } i^{de} \text{ kandidaat in de steekproef op kandidaat A stemt} \\ 0 \text{ als de } i^{de} \text{ kandidaat in de steekproef op kandidaat B stemt} \end{cases}$. Omdat $N \gg n$ mogen we veronderstellen dat X_i en X_j onafhankelijk van elkaar zijn als $i \neq j$.
$p_A = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dus geldt $Var(p_A) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}np(1-p)$ met $p = 0.53$. Er geldt dus:
$Var(P_A) = \frac{p(1-p)}{n} = \frac{0.53 \cdot 0.47}{1000} = \frac{(0.5+0.03)(0.5-0.03)}{1000} = \frac{(0.5^2-0.03^2)}{1000} = \frac{0.025-0.0009}{1000} = 0.0002491$.

b 7pt) Bepaal $Cov(p_A, p_B)$.

ANTWOORD: $p_A + p_B = 1$, dus $Var(p_A + p_B) = 0$. Er geldt $Var(p_A + p_B) = Var(p_A) + Var(p_B) + 2Cov(p_A, p_B)$. $Var(p_B) = Var(p_A)$ (herhaal de redenering van antwoord A), dus geldt $Cov(p_A, p_B) = -Var(p_A) = -0.0002491$.

Alternatief, je kunt de covariantie ook direct uitrekenen.
$Cov(p_A, p_B) = E(p_A - E(p_A)) \cdot E(p_B - E(p_B)) = E(p_A - 0.53)E(p_B - 0.47)$.
Omdat $p_A + p_B = 1$ geldt: $Cov(p_A, p_B) = E(p_A - 0.53)E(1 - p_A - 0.47) = E(p_A - 0.53)E(0.53 - p_A) = -E((p_A - 0.53)^2) = -E((p_A - E(p_A))^2) = -Var(p_A) = -0.0002491$

Opgave 2 X en Y zijn discrete stochastische variabelen op de uitkomstenruimte $(\mathbb{Z}^+)^2$, met $\mathbb{Z}^+$ de verzameling van alle positieve gehele getallen, i.e., $\mathbb{Z}^+ = \{1, 2, 3, \ldots\}$. De gezamenlijke kansfunctie van X en Y is gegeven door:

$$P(X = i, Y = j) = \left(\frac{1}{2}\right)^i i \left(\frac{1}{i+1}\right)^j$$

U mag in deze opgave gebruiken dat $\sum_{i=1}^{\infty} \frac{i}{i+1}(\frac{1}{2})^i = 2 - \log 4$.

a 8pt) Bepaal de marginale kansfunctie voor X.

ANTWOORD: $P(X = i) = \sum_{j=1}^{\infty} P(X = i, Y = j) = \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^i i \left(\frac{1}{i+1}\right)^j$
Er geldt: $\sum_{j=0}^{\infty} (r)^j = \frac{1}{1-r}$ als $|r| < 1$, dus $\sum_{j=1}^{\infty} (r)^j = \frac{1}{1-r} - 1 = \frac{r}{1-r}$
Dus $\sum_{j=0}^{\infty} \left(\frac{1}{i+1}\right)^j = \frac{\frac{1}{i+1}}{1-\frac{1}{i+1}} = \frac{1}{i}$. Hieruit volgt:
$P(X = i) = \left(\frac{1}{2}\right)^i i \frac{1}{i} = \left(\frac{1}{2}\right)^i$.

b 7pt) Zijn X en Y onafhankelijk?

ANTWOORD: Nee, $P(X=i, Y=j) \neq P(X=i)P(Y=j)$ voor alle i, j. Een tegenvoorbeeld is het geval $i=1$ en $j=1$.

$P(X=1, Y=1) = \frac{1}{2} 1 \left(\frac{1}{2}\right)^1 = \frac{1}{4}$.

$P(X=1) = \frac{1}{2}$

$P(Y=1) = \sum_{i=1}^{\infty} (\frac{1}{2})^i i (\frac{1}{i+1})^1 = 2 - \log 4$

Opgave 3 Een experimentator wil graag de concentratie C van stof A in een vloeistof bepalen. Daartoe neemt hij 4 monsters van de vloeistof en voegt aan elk monster een reagens toe die de vloeistof donkerder maken, hoe hoger de concentratie van stof A, des de donkerder de vloeistof wordt. Bij alle 4 de monsters bepaalt hij welk percentage van laserlicht door de verdunning komt. De experimentator weet dat de theoretische relatie tussen de concentratie C in het monster en het percentage doorgelaten licht I gegeven is door:

$$I = 100e^{-\frac{C}{2}}$$

De experimentator veronderstelt dat er bij elke meting van het percentage een standaard-normaal-verdeelde meetfout t.o.v. van de theoretische waarde, en dat alle metingen onafhankelijk van elkaar zijn. De gemeten percentages doorgelaten licht zijn: $34\%, 35\%, 36\%, 31\%$.

a 10pt) Bepaal de meest waarschijnlijke schatter van de concentratie C.

ANTWOORD: Definieer $x_1 = 34$, $x_2 = 35$, $x_3 = 36$ en $x_4 = 31$.

$L = \prod_{i=1}^{4} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - 100e^{-\frac{C}{2}})^2}{2}}$ en de loglikelihood is

$l = -2\log \pi - \sum_{i=1}^{4} \frac{(x_i - 100e^{-\frac{C}{2}})^2}{2}$.

$\frac{d}{dC} l = -\sum_{i=1}^{4} (x_i - 100e^{-\frac{C}{2}}) 50e^{-\frac{C}{2}} = 0$. Substitueer $w = 50e^{-\frac{C}{2}}$. Dan krijgen we: $\sum_{i=1}^{4} (x_i - 2w)w = 0$. Dit geeft $w = 0$, wat niet kan, of $\sum_{i=1}^{4} (x_i - 2w) = 0$. Invullen geeft: $136 - 8w = 0$, oftewel, $w = 17$, oftewel, $c = -2\log \frac{17}{50} (\approx 2.16)$

Opgave 4 Zij X en Y twee onafhankelijke uniform verdeelde stochasten op $(0,1)$. Definieer de stochast Z door $Z = -\frac{\log Y}{X}$.

a 10pt) Bepaal $P(Z > 1)$.

ANTWOORD: $P(Z > 1) = P(-\frac{\log Y}{X} > 1) = P(-\log Y > X) = \int_0^1 dy \int_0^{\min(-\log y, 1)} dx = \int_0^{e^{-1}} dy \int_0^1 dx + \int_{e^{-1}}^1 dy \int_0^{-\log y} dx = e^{-1} - \int_{e^{-1}}^1 dy \log y = e^{-1} - (y \log y - y)|_{e^{-1}}^1 = e^{-1} + 1 + (-e^{-1} - e^{-1}) = 1 - e^{-1}$.

Opgave 5 Stel dat het jaarinkomen van wiskundigen normaal verdeeld is volgens $N(\mu, \sigma^2)$. Een willekeurige steekproef onder 25 wiskundigen geeft als steekproefgemiddelde van het jaarinkomen €50.000,- en S_{25}=€5.000,-.

a 9pt Bepaal de waarde van het jaarinkomen x waarvoor geldt dat we 95% zekerheid weten dat $\mu > x$.
ANTWOORD: $t_x = \frac{50.000-x}{5.000/\sqrt{25}} = \frac{50.000-x}{1000}$

Voor een student-t-verdeling met 24 vrijheidsgraden is de kans dat de stochast groter is dan 1,711 precies 95% volgens Tabel B.2. We moeten dus oplossen: $t_x = 1,318 \Rightarrow \frac{50.000-x}{1000} = 1,711$. Dit geeft $x = 48.289$.

b 6pt Bepaal een tweezijdig 90%-betrouwbaarheidsinterval voor μ.
ANTWOORD: $(50000 - 1,711\frac{5000}{sqrt25}; 50000 + 1,711\frac{5000}{sqrt25}) = (48289; 51711)$.

Opgave 6 Zij Y het aantal auto's in een willekeurig gekozen huishouden. De kansfunctie van Y is gegeven in onderstaande tabel:

Y	0	1	2	3	4
P(Y)	0.25	0.44	0.2	0.08	0.03

Jan wil wedden dat als hij een willekeurige steekproef van grootte 225 uitvoert, dat het aantal auto's groter is dan 225. Hij is zo zeker van zijn zaak dat hij €1,- wint als hij gelijk heeft en €100,- verliest als er 225 of minder auto's worden geteld.

a 9pt Wat is de kans dat Jan deze weddenschap wint? Je mag ervan uitgaan dat de steekproefgrootte groot genoeg is om asymptotische benaderingen te mogen gebruiken.
ANTWOORD: De verwachtingswaarde van Y is $0.25 \cdot 0 + 0.44 \cdot 1 + 0.2 \cdot 2 + 0.08 \cdot 3 + 0.03 \cdot 4 = 0 + 0.44 + 0.4 + 0.24 + 0.12 = 1.2$. De variantie van Y is $Var(Y) = E(Y^2) - E(Y)^2 = 0.25 \cdot 0^2 + 0.44 \cdot 1^2 + 0.2 \cdot 2^2 + 0.08 \cdot 3^2 + 0.03 \cdot 4^2 - 1.2^2 = 0.44 + 0.8 + 0.72 + 0.48 - 1.44 = 2.44 - 1.44 = 1$.

Volgens de centrale limietstelling is het steekproefgemiddelde bij benadering normaal verdeeld als $N(1.2, \frac{1}{225})$. De kans die we moeten uitrekenen is $P(aantal \quad autos > 225) = P(\bar{Y}_{225} > 1) = P(\bar{Y}_{225} - 1.2 > -0.2) = P(\frac{\bar{Y}_{225}-1.2}{\frac{1}{15}} > -0.2 \cdot 15) = P(\frac{\bar{Y}_{225}-1.2}{\frac{1}{15}} > -3) \approx P(N(0,1) > -3) = 1 - 0.0013 = 0.9987$.

b 6pt Wat is de verwachtingswaarde van de weddenschap? Als je a) niet hebt beantwoord, neem dan aan dat de kans op het winnen van de weddenschap 0.85 is.
ANTWOORD: $0.9987 \cdot 1 + 0.0013 \cdot (-100) = 0.8687$.

Opgave 7 Zij A een verzameling van n verschillende reële getallen, i.e., $A = \{x_1, x_2, \ldots x_n\}$. Zij $x_{(i)}$ het element van A zodanig dat er precies $(i-1)$ elementen van A kleiner zijn dan $x_{(i)}$, $x_{(1)}$ is dus het kleinste element van A en $x_{(n)}$ het grootste element van A. Stel dat de elementen van A oorspronkelijk in een willekeurige volgorde geordend zijn waarbij elke volgorde even waarschijnlijk is.

a 4pt) Hoeveel mogelijke ordeningen van de verzameling A bestaan er?
ANTWOORD: n elementen kunnen op $n!$ manieren geordend worden.

b 6pt) Bubblesort is een algoritme om n elementen op volgorde te leggen, zodanig dat het het kleinste element op positie 1 is, het een-na-kleinste element op positie 2, ... , en het grootste element op positie n. Bubble-sort kijkt naar naast-elkaar gelegen elementen en verwisselt ze van volgorde als het element op positie i groter is dan het element op positie $i+1$. Het algoritme vergelijkt (en verwisselt indien nodig) eerst de elementen op positie 1 en 2, dan op positie 2 en 3, tot op positie $n-1$ en n. Dan weet je zeker dat het grootste element op positie n is. Vervolgens start je opnieuw en vergelijk je (en verwissel je indien nodig) eerst de elementen op positie 1 en 2, dan op positie 2 en 3, tot op positie $n-2$ en $n-1$. Nu ligt ook het element $x_{(n-1)}$ op positie $n-1$. Deze procedure wordt herhaald tot alle elementen op volgorde liggen.

Voor deze vraag zijn we geïnteresseerd in het aantal keren k dat elementen verwisseld worden. Het aantal verwisselingen is gelijk aan het aantal paren (i,j), met $i<j$, in de oorspronkelijke configuratie waarvoor $x_i > x_j$.

Definieer $P_n(k)$ als de kans dat je k keer 2 elementen van positie moet wisselen voordat de n elementen op volgorde liggen.

Geef de verwachtingswaarde van het aantal verwisselingen.
ANTWOORD: Er zijn $\frac{n(n-1)}{2}$ paren. De kans dat een willekeurig gekozen paar in de juiste volgorde ligt is 1/2. De verwachtingswaarde is lineair, dus $E(k) = \frac{n(n-1)}{2}\frac{1}{2} = \frac{n(n-1)}{4}$.

c 10pt) Bonusvraag: Bepaal de variantie van het aantal verwisselingen.
ANTWOORD: $k = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathbb{1}_{x_i > x_j} = \sum_{paren} \mathbb{1}_{paar\ niet\ geordend}$.
$Var(k) = E(k^2) - E(k)^2$
$k^2 = \sum_{i=1}^{\#pairs}\sum_{j=1}^{\#pairs} \mathbb{1}_{\text{paar } i \text{ niet geordend}}\mathbb{1}_{\text{paar } j \text{ niet geordend}}$
$E(k^2) = \sum_{i=1}^{\#pairs}\sum_{j=1}^{\#pairs} E(\mathbb{1}_{\text{paar } i \text{ niet geordend}}\mathbb{1}_{\text{paar } j \text{ niet geordend}})$.
We gaan nu kijken naar paren van paren.

Er zijn $n(n-1)/2$ identieke paren, dus van de vorm $(x_i,x_j)(x_i,x_j)$ met $i<j$. De helft van deze paren is ongeordend.

Beschouw nu paren zonder gemeenschappelijke elementen, dus van de vorm $(x_i,x_j)(x_k,x_l)$ met $i<j$, $k<l$ en i, j, k en l allemaal verschillende. Uit n elementen kun je $n(n-1)/2$ paren maken. Als je het eerste paar hebt gekozen heb je nog $(n-2)$ elementen over om uit te kiezen, dus je hebt dan nog $(n-2)(n-3)/2$ opties voor het tweede paar. In totaal heb je dus $n(n-1)(n-2)(n-3)/4$ paren zonder gemeenschappelijke elementen. Elk paar heeft 50% kans om ongeordend te zijn, dus in een kwart van de gevallen zijn beide paren ongeordend.

We gaan nu naar de paren met 1 gemeenschappelijk element. Hiervoor zijn er 4 opties, met $i < j$, $i \neq j$, $i \neq k$ en $j \neq k$.

1. $(x_i, x_j)(x_k, x_i)$ met $k < i$
2. $(x_i, x_j)(x_j, x_k)$ met $j < k$
3. $(x_i, x_j)(x_k, x_j)$ met $k < j$
4. $(x_i, x_j)(x_i, x_k)$ met $i < k$

- Het aantal paren van paren van type 1 is $\frac{n(n-1)(n-2)}{6}$. $n(n-1)(n-2)$ is het aantal manieren waarop je 3 verschillende elementen uit een verzameling van n elementen kunt kiezen. In één van de zes gevallen geldt dan dat $k < i < j$. Dit kun je eventueel ook direct uitrekenen $\sum_{i=2}^{n-1} \sum_{j=i+1}^{n} \sum_{k=1}^{i-1} 1 = \frac{n(n-1)(n-2)}{6}$.

  Alleen als $x_k > x_i > x_j$ zijn beide paren ongeordend, de kans hierop is $1/6$.
- Het aantal paren van type 2 is ook $\frac{n(n-1)(n-2)}{6}$. $n(n-1)(n-2)$ is het aantal manieren waarop je 3 verschillende elementen uit een verzameling van n elementen kunt kiezen. In één van de zes gevallen geldt dan dat $i < j < k$.

  Alleen als $x_i > x_j > x_k$ zijn beide paren ongeordend, de kans hierop is $1/6$.
- Het aantal paren van type 3 is $\frac{n(n-1)(n-2)}{3}$. $n(n-1)(n-2)$ is het aantal manieren waarop je 3 verschillende elementen uit een verzameling van n elementen kunt kiezen. In twee van de zes gevallen geldt dat $i < j$ en $k < j$.

  Alleen als $x_i > x_j$ en $x_k > x_j$ geldt dat beide paren ongeordend zijn. De kans hierop is $1/3$.
- Het aantal paren van type 4 is ook $\frac{n(n-1)(n-2)}{3}$. $n(n-1)(n-2)$ is het aantal manieren waarop je 3 verschillende elementen uit een verzameling van n elementen kunt kiezen. In twee van de zes gevallen geldt dat $i < j$ en $i < k$.

  Alleen als $x_i > x_j$ en $x_i > x_k$ geldt dat beide paren ongeordend zijn. De kans hierop is $1/3$.

Hieruit volgt dat $E(k^2) = \frac{n(n-1)}{2}\frac{1}{2} + \frac{n(n-1)(n-2)(n-3)}{4}\frac{1}{4} + \frac{n(n-1)(n-2)}{6}\frac{1}{6} + \frac{n(n-1)(n-2)}{6}\frac{1}{6} + \frac{n(n-1)(n-2)}{3}\frac{1}{3} + \frac{n(n-1)(n-2)}{3}\frac{1}{3} = \frac{n(n-1)}{4} + \frac{n(n-1)(n-2)(n-3)}{16} + \frac{10n(n-1)(n-2)}{36} = \frac{n(n-1)}{4}(1 + \frac{(n-2)(n-3)}{4} + \frac{10}{9}(n-2)) = \frac{n(n-1)}{144}(36 + 9(n-2)(n-3) + 40(n-2)) = \frac{n(n-1)}{144}(9n^2 - 5n + 10)$.

Dus geldt $Var(k) = E(k^2) - E(k)^2 = \frac{n(n-1)}{144}(9n^2 - 5n + 10) - \left(\frac{n(n-1)}{4}\right)^2 = n(n-1)\left(\frac{9}{144}n^2 - \frac{5}{144}n + \frac{10}{144} - \frac{1}{16}n^2 + \frac{1}{16}n\right) = n(n-1)\left(\frac{5}{144}n + \frac{10}{144}\right) = \frac{n(n-1)(2n-5)}{72}$.

Einde.