Examen Kansrekening en Wiskundige Statistiek

S. Vansteelandt

Academiejaar 2005-2006

---

1. Stel dat $X_1, X_2, ...$ een oneindige reeks onafhankelijke en identiek verdeelde toevalsveranderlijken zijn met gemiddelde 2 en variantie 1. Stel dat we onafhankelijk van deze waarden een lukrake waarde Y trekken uit een Poisson verdeling met gemiddelde 10, en vervolgens de waarden X_1 tot X_Y optellen. Wat is dan de

   (a) verwachtingswaarde

   (b) variantie

   van de resulterende som $Z = \sum_{i=1}^{Y} X_i$?

2. Geef aan of de volgende uitspraken JUIST of FOUT zijn. Leg telkens uit waarom, ongeacht of de uitspraak juist of fout is.

   (a) De correlatie tussen de diastolische bloeddruk en body mass index bedraagt 0.5. Dit betekent dat mensen met hoge BMI doorgaans ook een hogere bloeddruk hebben.

   (b) Een statistische toets geeft op basis van 18 observaties aan dat er geen significante associatie is tussen hypertensie en het risico om complicaties op te doen na het plaatsen van een stent (p-waarde 0.32). We kunnen hieruit besluiten dat patiënten met hoge of lage bloeddruk evenveel risico lopen op complicaties.

3. De correctheid van het verpakkingsmechanisme van een graanhandelaar staat in vraag. Het verpakkingsmechanisme is correct als het echte gemiddelde van een zak graan 12 kg of meer bedraagt. De handelaar beslist een eenvoudige lukrake steekproef te nemen om de machine te controleren. Hij zal de machine behouden als het gemiddelde $\bar{X}$ groot genoeg is anders niet.

   (a) Specifieer het beslissingscriterium en de steekproefgrootte als de handelaar niet meer dan 5% risico wil lopen om de machine te vervangen als het ware gemiddelde toch voldoet en hij ten hoogste 10% risico wil lopen om de machine toch te behouden als het gemiddelde slechts 11.6 kg bedraagt. Men weet uit ervaring dat het aantal kilogram in 1 zak Normaal verdeeld is met een standaardafwijking van 0.5 kg.

   (b) Wat is de lengte van het 95% betrouwbaarheidsinterval voor het gemiddelde gewicht op basis van een steekproef van de hierboven berekende grootte?

4. Zij $X_1, ..., X_n$ onafhankelijke toevalsveranderlijken die Normaal verdeeld zijn met een zelfde gemiddelde θ en even grote variantie θ; d.i. $E(X_i) = \text{Var}(X_i) = \theta$ voor $i = 1, ..., n$.

   (a) Bereken dan de maximum kans schatter $\hat{\theta}$ voor θ (oplossing $\hat{\theta} = \sqrt{0.25 + \overline{x^2}} - 1/2$).

   (b) Toon aan dat deze schatter consistent is.

   (c) Toon aan dat deze schatter asymptotisch Normaal verdeeld is.

5. Een onderzoek werd ingericht om aan te tonen dat sojabonen ingeënt met nitrogeen-vasthoudende bacteriën meer opbrengen en adequaat groeien zonder het gebruik van dure voor het milieu schadelijke synthetische meststof. Het onderzoek werd uitgevoerd onder gecontroleerde condities met uniforme hoeveelheden grond. Er werden 16 planten beschouwd, waarvan 8 ingeënt en niet ingeënt. De plantenopbrengst werd gemeten door het peulengewicht (gram) van iedere plant en er werd telkens een gemiddelde en standaarddeviatie berekend:

	ingeënt	niet ingeënt	verschil
	1.76	0.49	1.27
	1.45	0.85	0.60
	1.03	1.00	0.03
	1.53	1.54	−0.01
	2.34	1.01	1.33
	1.96	0.75	1.21
	1.79	2.11	−0.32
	1.21	0.92	0.29
gem	1.634	1.084	0.550
sd	0.420	0.510	0.652

   (a) Geef een p-waarde bij de toets of de inenting met bacteriën effect heeft op de gemiddelde plantenopbrengst.

   (b) Rapporteer een zinvolle maat om het effect van inoculatie op de plantenopbrengst uit te drukken. Bereken tevens een bijhorend **99%** betrouwbaarheidsinterval.

   (c) Trek op basis van uw antwoord op vragen (a) en (b) een zorgvuldig en concreet kwantitatief[1] besluit omtrent het verschil tussen beide methodes met betrekking tot het effect van inenting op de plantenopbrengst.

   (d) Aan welke voorwaarde(n) moet voldaan zijn opdat het betrouwbaarheidsinterval uit (b) de bijhorende populatieparameter met 99% kans zou omvatten? Leg uit.

6. Veronderstel dat $X_n, n = 1, 2, ...$ een reeks Poisson verdeelde toevalsveranderlijken zijn met gemiddelde n.

   (a) Ga dan na naar welke verdeling
   $$\frac{X_n - n}{\sqrt{n}}$$
   convergeert naarmate $n \to \infty$.

   (b) Bovenstaand resultaat geeft meer algemeen aan naar welke verdeling Poisson toevalsveranderlijken convergeren naarmate hun gemiddelde λ groot wordt. Gebruik het om een (benaderende) formule op te stellen voor het 95% betrouwbaarheidsinterval voor de verwachtingswaarde λ van een Poisson verdeeld veranderlijke bij grote waarden van λ.

Veel succes!

[1] M.a.w. bespreek niet alleen of inenting al dan niet beter is, maar ook de grootte van het effect.

Statistiek I: Examen 2e zit 2005-2006

Peter Vandendriessche

16 januari 2007

1. (a) $E(Z) = E(E(Z|Y)) = E(2Y) = 20$
   (b) $\mathrm{Var}(Z) = E(\mathrm{Var}(Z|Y)) + \mathrm{Var}(E(Z|Y)) = E(Y) + \mathrm{Var}(2Y) = 50$
2. (a) Juist, duidelijk positieve correlatie impliceert dat verband.
   (b) Fout, je kunt enkel niet besluiten dat het wel zo zou zijn.
3. (a) We doen hier een one-sample z-test. We zullen dus het stelsel
$$\left\{ \frac{\bar{X} - 11.6}{0.5/\sqrt{n}} = z_{0.90}, \frac{\bar{X} - 12}{0.5/\sqrt{n}} = z_{0.05} \right\}$$
   oplossen. Dit geeft $\sqrt{n} = \frac{0.5\cdot(1.282+1.645)}{0.4}$ ofte $n = 13.386$. De grootte is dus minstens 14.
   (b) Het 95% BI is $\left[\bar{X} - z_{0.975}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{0.975}\frac{\sigma}{\sqrt{n}}\right]$ dus de lengte is $2 \cdot \frac{\sigma}{\sqrt{n}} = 0.524$.
4. (a) We hebben $L(\theta) = \frac{1}{\sqrt{2\pi\theta}^n} e^{-\frac{1}{2}\frac{\sum(x_i-\theta)^2}{\theta}}$, dus $\ell' = -\frac{n}{2\theta} + \frac{\sum x_i^2}{2\theta^2} - \frac{n}{2} = 0$ geeft de VKV $\frac{\sum x_i^2}{n} - \theta - \theta^2 = 0$ zodat $\theta = \frac{-1+\sqrt{1+4\overline{x^2}}}{2}$.
   (b) $E(\overline{x^2}) = \mathrm{Var}(x_i) + E(x_i)^2 = \theta + \theta^2$ zodat $t_n \xrightarrow{p} \sqrt{\theta^2 + \theta + \frac{1}{4}} - \frac{1}{2} = \theta$.
   (c) CLT zegt ons dat $\sqrt{n}(\overline{x^2} - (\theta^2 + \theta)) \xrightarrow{d} N(0, Var)$, zodat de delta-methode met $g(x) = \sqrt{\frac{1}{4} + x} - \frac{1}{2}$ ons geeft dat $\sqrt{n}(t_n - \theta) \xrightarrow{d} N\left(0, \frac{Var}{1+4\theta}\right)$, dus asymptotisch normaal.
5. (a) We doen een ongepaarde t-test. Onderstel dat beide populaties normaal verdeeld zijn met zelfde variantie. We hebben dat $S_P = \frac{7\cdot 0.42^2 + 7\cdot 0.51^2}{14} = 0.218$, zodat $T = \frac{\bar{X}_8 - \bar{Y}_8}{S_P/\sqrt{\frac{1}{4}}} = 2.355 \in [t_{7,0.975}, t_{7,0.99}]$. Dus ligt de p-waarde tussen 0.02 en 0.05.
   (b) Als maat nemen we het gemiddeld verschil tussen beide groepen. Een 99% BI voor het verschil is $\left[0.550 - t_{14,0.995} S_P \sqrt{\frac{1}{4}}, 0.550 + t_{14,0.995} S_P \sqrt{\frac{1}{4}}\right] = [-0.145, 1.245]$.
   (c) Met 99% kans ligt het werkelijke gemiddeld verschil in dit interval. Het interval bevat 0, dus we verwerpen de hypothese dat er geen verschil is niet op 99% significantie. De p-waarde zegt ons echter dat we die hypothese op 95% significantie wel mogen verwerpen.
   (d) Zoals eerder vermeld onderstellen we dat beide populaties normaal verdeeld zijn met dezelfde variantie.
6. (a) $X_n = \sum_{i=1}^n Y_i$ met $Y_i \stackrel{d}{=}$, zodat wegens CLT $\frac{X_n - n}{\sqrt{n}} = \sqrt{n}(\bar{Y}_n - 1) \xrightarrow{d} N(0,1)$.
   (b) $Po(\lambda) \stackrel{d}{\approx} N(\lambda, \lambda)$ zodat dit interval $[\lambda - z_{0.975}\sqrt{\lambda}, \lambda + z_{0.975}\sqrt{\lambda}]$ is.