Namn: Lucas Frykman

Personnummer: 0210127650

Kurskod: SF1930

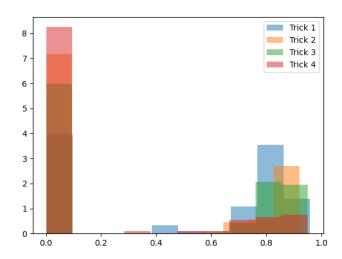
Kursansvarig:: Liam Solus



# Statistisk inlärning och dataanalys Projekt October 15, 2023

# 1 Uppvärmning

Figure 1: Histogram av betyg skalad mellan 0 och 1



Låt B vara betyg för en ett skateboardåkare och trick. Vi vill skatta  $P(B>0.6|B>0)=\frac{P(B>0|B>0.6)P(B>0.6)}{P(B>0)}=\frac{P(B>0.6)P(B>0.6)}{P(B>0)}$  som  $\tilde{P}(B>0.6|B>0)=\frac{\sum_{i}^{4}\sum_{j}^{96}trick_{ij}\mathbf{1}_{\{[0.6,1]\}}}{\sum_{i}^{4}\sum_{j}^{96}trick_{ij}\mathbf{1}_{\{[0,1]\}}}\approx 0.96$  Det här stämmer med utseendet på fig. 1. När man plottar run 2 mot run 1 ser de ut att ha jätte svag korrelation fig. 2

#### 2 En frekventistisk modell

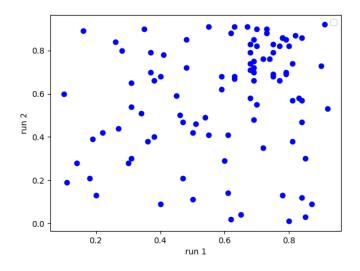
Anm 1 Vår model för  $X_i$  är följande  $X_i = \begin{cases} 0 \text{ om } V_i = 0 \\ Z_i \text{ om } V_i = 1 \end{cases}$ 

 $d\ddot{a}r\ V_i \sim Ber(\theta_i)\ och\ Z_i \sim Beta(\alpha_i,\beta_i)\ det\ h\ddot{a}r\ \ddot{a}r\ ekvivalent\ med\ att\ s\ddot{a}ga$ 

 $V_i = \mathbf{1}_{\{x \neq 0\}}(X_i) \text{ och } Z_i = X_i | (V_i = 1)$ 

eftersom det här är bara en transformation av stokatiska variebler ger stickprov från  $X_i$  oss ett stickprov för  $Z_i$  och  $V_i$ 

Figure 2: Spridningsdiagram mellan run 1 och run 2



### (a) Skatta $\theta_i$

tricket). Nu tar vi log likliehoodfunktionen.

Låt  $x_{i[n]} = (x_{i1}, x_{i2}, ... x_{in})^T$  vara vår stickprov från samtliga trick skateboardåkaren i utförde.

$$L(\theta_i, \alpha_i, \beta_i | x_{i[n]}) = \prod_{j=1}^n f_{x_i}(x_{ij}) = \prod_j^n (1 - \theta_i) \mathbf{1}_{\{x=0\}}(x_{ij}) + \theta_i f_{Z_i}(x_{ij}) \mathbf{1}_{\{x \neq 0\}}(x_{ij})$$
(1)

$$\{x=0\}(x_{ij})$$
 (2)

 $L(\theta_i, \alpha_i, \beta_i | x_{i[n]}) = (1 - \theta_i)^{n-m} \theta_i^m \prod_{i=1}^n \left( f_{Z_i}(x_{ij}) \mathbf{1}_{\{x \neq 0\}}(x_{ij}) + \mathbf{1}_{\{x=0\}}(x_{ij}) \right)$ där  $m = \sum_{j=1}^{n} \mathbf{1}_{\{x \neq 0\}}(x_{ij})$  alltså hur många gånger  $x_i$  inte är noll (gånger tävlaren i landade

$$\implies \log(L) = (n - m)\log(1 - \theta_i) + m\log(\theta_i) + \sum_{i=1}^{n} \log(f_{Z_i}(x_{ij})\mathbf{1}_{\{x \neq 0\}}(x_{ij}) + \mathbf{1}_{\{x = 0\}}(x_{ij}))$$
(3)

$$\iff \partial_{\theta_i} \log(L) = \frac{m-n}{1-\theta_i} + \frac{m}{\theta_i} = 0$$
 (4)

$$\iff \frac{m - n\theta_i}{\theta_i (1 - \theta_i)} = 0 \iff \hat{\theta_i} = \frac{m}{n}$$
 (5)

MLE för bernoulli fördelningens  $V_i$  parameter  $\hat{\theta_i} = \mathrm{argmax}_{\theta \in \Omega} L(\theta_i | v_{i[n]}) = \bar{v_i}$  skulle ge oss samma resultat. Eftersom vi kan transformera stickprovet  $x_{i[n]} \to v_{i[n]}$  med anm 1  $v_i = \mathbf{1}_{\{x \neq 0\}}(x_i)$ . vilket betyder att  $m = \sum_{j=1}^{n} v_i$  och därmed får eq. (5) att sammanfalla med MLE av bernoulli fördelningen.

### (b) skatta $\alpha_i$ och $\beta_i$

Observera att från eq. (3)  $\sum_{j=1}^{n} \log \left( f_{Z_i}(x_{ij}) \mathbf{1}_{\{x \neq 0\}}(x_{ij}) + \mathbf{1}_{\{x=0\}}(x_{ij}) \right) = \sum_{j=1}^{n} \log \left( f_{Z_i}(x_{ij}) \mathbf{1}_{\{x \neq 0\}}(x_{ij}) \right)$  eftersom  $\log(1) = 0$ . Vi vet att  $\arg\max_{\alpha,\beta\in\Omega} \log(L) = \arg\max_{\alpha,\beta\in\Omega} \sum_{j=1}^{n} \log \left( f_{Z_i}(x_{ij}) \mathbf{1}_{\{x \neq 0\}}(x_{ij}) \right)$  vilket är ekvivalent med  $\arg\max_{\alpha,\beta\in\Omega} \log(L(\alpha,\beta|z_{i[k]}))$  för att z stickprovet innehåller alla trick som landade  $z_{i[k]} = (z_{i1}, \dots z_{ik})^T = \{x_{ij} \in x_{i[n]} : x_{ij} \neq 0\}$ 

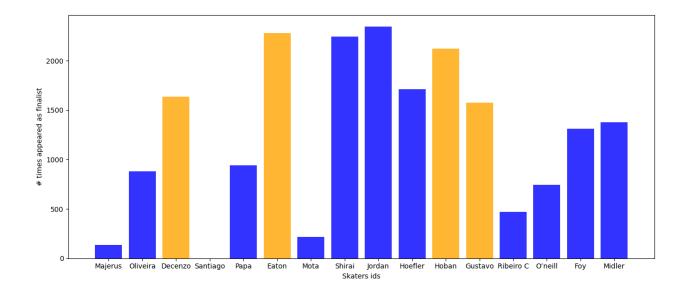
Vi ska alltså bara maximera log-likelihood av beta fördelningens paramtrerna givet data from  $Z_i$ 

$$\begin{cases} \partial_{\alpha} \log(L(\alpha, \beta | z_{i[k]})) = \sum_{j=1}^{k} \partial_{\alpha} \log(f(z_{ij})) = 0 \\ \partial_{\beta} \log(L(\alpha, \beta | z_{i[k]})) = \sum_{j=1}^{k} \partial_{\beta} \log(f(z_{ij})) = 0 \end{cases} \\ :: \partial_{\alpha} \log(f(z_{ij})) = \partial_{\alpha} \log \left( \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot z_{ij}^{\alpha - 1} \cdot (1 - z_{ij})^{\beta - 1} \right) \\ = \partial_{\alpha} \left( \log \Gamma(\alpha + \beta) - \log \Gamma(\alpha) - \log \Gamma(\beta) + (\alpha - 1) \log z_{ij} + (\beta - 1) \log(1 - z_{ij}) \right) \\ = \partial_{\alpha} \log(f(z_{ij})) = \psi(\alpha + \beta) - \psi(\alpha) + \log z_{ij} \operatorname{där} \psi = \Gamma' / \Gamma \\ (\operatorname{Vi \ g\"{o}r \ liknande \ f\"{o}r \ } \partial_{\beta} \log f(z_{ij})) \\ \Rightarrow \begin{cases} \partial_{\alpha} \log L = k\psi(\alpha + \beta) - k\psi(\alpha) + \sum_{j=1}^{k} \log(z_{ij}) = 0 \\ \partial_{\beta} \log L = k\psi(\alpha + \beta) - k\psi(\beta) + \sum_{j=1}^{k} \log(1 - z_{ij})) = 0 \end{cases}$$

Det går dock inte att lösa ML skattningen analytisk härifrån. Numeriska metoder som newton rhapson eller gradient descent behövs för att skatta vår ML skattning. Att göra så medför sig en del problem som ökar systematiska felet genom numerisk fel. Vi behåller riktighet i punktskattningen genom att använda moment metoden istället. Vi utgår från följande system ekvationerna

För vissa skatebordåkare kan man inte få en punktskattningen eftersom det finns endast en datapunkt  $z_1$  för vissa skatebordåkare. Stickprovsvariansen i sådana fall blir lika med noll. Stickprovsvariansen representerar hur osäkert vi är om en skatebordåkares prestation på en trick. Intuivt sett Vi gör valet här att skatta stickprovsvariansen som  $S_i^2 \approx \bar{S}^2$  dvs vi tar medelvärdet av samtliga varianser.

Figure 3: Frequency appearing in final W



### (c) Model för $Y_i$

 $Y_i$  ska vara run betyget för åkaren i. Eftersom  $Y_i \in (0,1]$  (varje deltagare får betyg större än 0) kommer bernoulli delen försvinna. Men annars fungerar en run ganska liknande som en trick. Så vi antar att  $Y_i \sim \text{Beta}({}^y\alpha_i, {}^y\beta_i)$ . För att skatta parameter till Y använder vi samma moment metoden som vi använde för att skatta  $\alpha, \beta$  för X. (Obs: Enda skillnaden i det här fallet vi slipper använda medelvärdet på stickprovsvariansen  $S^2 = \bar{S}^2$  för att det är garanterat vi får mer än en datapunkt för alla skateboardåkare i.)

#### (d) LCQ simularing för frekventistisk modell

Total betyget för varje deltagare beräknas som summan av deras två största trick betyg och största run betyg. Vi kan beskriva i termer av stokatiska variabler. Låt  $O_i$  vara total betyg för deltagare i. Låt  $Q_{i,\text{först}} = \max(X_{i1}, X_{i2}, X_{i3}, X_{i4})$  och

 $Q_{i,\text{andra}} = \max(\min(X_{i1}, X_{i2}), \min(X_{i1}, X_{i3}), \min(X_{i1}, X_{i4}), \min(X_{i2}, X_{i3}), \min(X_{i2}, X_{i4}), \min(X_{i3}, X_{i4})).$  Vi vill simulera total betyget för varje deltagare i som  $O_i = Q_{i,\text{först}} + Q_{i,\text{andra}} + \max(Y_{i1}, Y_{i2})$  De som fick de 4 högsta betyg får delta i finalen. Vi simulerar 5000 LCQ:ar. Det ger oss en följd av stokatisk sets  $\mathbf{W}_1, \dots \mathbf{W}_{5000}$ . Python har redan libraries för att generera stickprov från beta och bernoulli fördelningar. Så vi slipper använda box muller, inverse metoden, eller dylikt. Jag skapade en frequency bar graph för att visualisera simuleringen och fick detta i en körning fig. 3. De som har markerats i orange är de som faktiskt vann den verkliga LCQ:en. Typvärdet på  $\mathbf{W}$  innehöll oftast Hoban, Eaton, Jordan, och Shirai. Typvärdets frekvens vara kring 50 gger dvs 1%. Frekvensen av när  $\mathbf{W}$  innehöll samtliga verkliga vinnare (nämligen Gustavo, Hoban, Eaton, Decenzo) vara när 16-20 gger dvs mindre än 0.5%.

# 3 En bayesiansk modell

### (a) Apriori fördelnignar

Vi ska föreslå en simultan apriori fördelningn för  $[\Theta_i, A_i, B_i]^T$ .

#### Apriori för $\Theta$

Vi vet att  $\theta$  representerar hur pass sannolikt det är att skatebordåkaren landar tricket och att det är ett värde mellan 0 och 1. Vi har dock inga starka åsikter om  $\theta$  innan vi observerar datan. Därför använder vi en icke-informativ apriori.

$$f_{\theta_i}(\theta_i) \propto 1$$

#### Apriori för A, B

Vi har inte någon vettig anledning att anta  $A \perp B$  så vi behöver en egentlig simultan fördelning som har stöd  $\alpha, \beta \in (0, \infty] \times (0, \infty]$ . För att uppnå detta omformulerar vi fördelningen med avseende på dess medelvärde  $\mu$  och en mått på precision  $\kappa$ . Det vill säga, för  $X \sim \text{Beta}(\alpha, \beta)$  tar vi

$$\mu = \frac{\alpha}{\alpha + \beta}$$
 och  $\kappa = \alpha + \beta + 1$ 

 $\kappa$  är ett mått på precision för fördelningen eftersom det är omvänt proportionellt mot variansen:

$$Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\mu(1-\mu)}{\kappa}.$$

Vi kan sedan ange en fördelning  $[\mu, \kappa]^T$  (enligt en hierarki) och sedan omvandla denna fördelning tillbaka till parametrarna  $\alpha$  och  $\beta$  för att få en fördelning för  $[A, B]^T$ . Vi tar den hierarkiska modellen:

$$\kappa | \mu \sim \text{Gamma}(\theta, \lambda), \mu \sim \text{U}(0, 1)$$

En variabelbyte ger oss den priorfördelning för parametrarna  $\alpha$  och  $\beta$ :

$$f_{A,B}(\alpha,\beta) = \frac{\lambda^{\theta}}{\Gamma(\theta)} (\alpha + \beta + 1)^{\theta - 1} e^{-\lambda(\alpha + \beta + 1)} (\alpha + \beta)^{-1}.$$

Vi antar  $\Theta_i \perp A_i, B_i$  för alla  $i \Longrightarrow f_{\theta_i,\alpha_i,\beta_i}(\theta_i,\alpha_i,\beta_i) = f_{\theta_i}(\theta_i) f_{\alpha_i,\beta_i}(\alpha_i,\beta_i)$ 

# (b) Aposteriori för $X_i$ och skattning

 $f_{\theta_i,\alpha_i,\beta_i|x_i}(\theta_i,\alpha_i,\beta_i|x_i) \propto f_{\theta_i}(\theta_i)f_{\alpha_i,\beta_i}(\alpha_i,\beta_i)f_{x_i|\theta_i,\alpha_i,\beta_i}(x_i|\theta_i,\alpha_i,\beta_i)$ Vårt mål är att använda aposteriorin för att skatta  $E[\Theta_i,A_i,B_i|\mathbf{X_i}=\mathbf{x_i}]$ . För att göra detta använder vi metropolis algoritimen för att generera ett stickprov och ta stickprovsmedelvärdet

$$\begin{pmatrix} \theta_{i0}, \theta_{i1} \dots \theta_{i5000} \\ \alpha_{i0}, \alpha_{i1} \dots \alpha_{i5000} \\ \beta_{i0}, \beta_{i1} \dots \beta_{i5000} \end{pmatrix} \longrightarrow \begin{pmatrix} \bar{\theta}_i \\ \bar{\alpha}_i \\ \bar{\beta}_i \end{pmatrix}$$

Aposteriorin är vår målfördelning men vi kommer använda  $L_p(\theta_i, \alpha_i, \beta_i) \propto \log(f_{\theta_i, \alpha_i, \beta_i | \mathbf{x}_i}(\theta_i, \alpha_i, \beta_i | \mathbf{x}_i))$  istället. Data fördelning  $f_{\mathbf{x}_i | \theta_i, \alpha_i, \beta_i}(\mathbf{x}_i | \theta_i, \alpha_i, \beta_i)$  i aposterorin är samma som likelihoodfunktionen  $L(\theta_i, \alpha_i, \beta_i | \mathbf{x}_i)$ 

$$\Rightarrow L_p(\theta_i, \alpha_i, \beta_i) = \underbrace{\log(f_{\theta_i})(\theta_i)}_{\log(1) = 0} + \log(f_{\alpha_i, \beta_i}(\alpha_i, \beta_i)) + \log L(\theta_i, \alpha_i, \beta_i)$$

#### Algorithm 1 Metropolis Algorithm

```
1: Initialize the chain with \Theta_0 = [\theta_{i0}, \beta_{i0}, \alpha_{i0}]^T
 2: for k = 1 to 5000 do
                                                                                    ▶ Iterate for a total of 5000 steps
         Generate a proposal state \Theta' from a proposal distribution Q(\Theta'|\Theta_{k-1})
         Calculate the acceptance ratio R = \min\left(1, \frac{L_p(\Theta')}{L_p(\Theta_{k-1})}\right), where L_p(\Theta) is the target distribu-
 4:
    tion
         Generate a uniform random number u \sim \text{Uniform}(0,1)
 5:
         if u \leq R then
 6:
             Accept the proposal: \Theta_k = \Theta'
 7:
 8:
 9:
             Reject the proposal: \Theta_k = \Theta_{k-1}
         end if
10:
11: end for
```

Valet för förslagsfördelning blir 
$$\mathbf{\Theta}' \sim Q(\mathbf{\Theta}' | \mathbf{\Theta}_{k-1}) = \begin{pmatrix} \exp(\log \alpha_{k-1} + \delta Z_{\alpha}) \\ \exp(\log \beta_{k-1} + \delta Z_{\beta}) \\ U_{\theta} \end{pmatrix} där Z_{\alpha} \sim N(0, 1),$$

 $Z_{\beta} \sim N(0,1)$ ,  $U_{\theta} \sim U(0,1)$ , och  $\delta = 0.5$  Anledning till valet för  $\alpha, \beta$  elementen är att det behöver vara ett tal större än 0 så vi tog exponentialen. Dessutom behöver vi  $\theta \in (0,1)$  så det var intuitivt att välja en uniform fördelning.

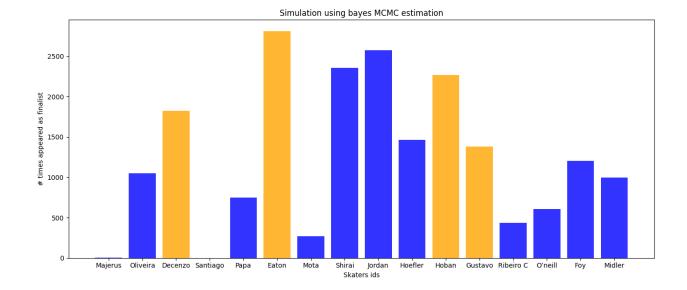


Figure 4: Frequency appearing in final W using MCMC estimated parameters

# (c) Aposteriori för $Y_i$

Eftersom  $Y_i \sim \text{Beta}({}^y\alpha_i, {}^y\beta_i)$  ska vi använda en apriori fördelning av samma form som den alpha beta apriori i X.

beta apriori i X.  $f_{\alpha_i,\beta_i}(\alpha_i,\beta_i) = \frac{\lambda^{\theta}}{\Gamma(\theta)}(\alpha_i + \beta_i + 1)^{\theta-1}e^{-\lambda(\alpha_i + \beta_i + 1)}(\alpha_i + \beta_i)^{-1} \text{ som ger oss aposteriorin}$ 

$$f_{\alpha_i,\beta_i|\mathbf{y}_i}(\alpha_i,\beta_i|\mathbf{y}_i) \propto f_{\alpha_i,\beta_i}(\alpha_i,\beta_i) f_{\mathbf{y}_i|\alpha_i,\beta_i}(\mathbf{y}_i|\alpha_i,\beta_i)$$

# (d) LCQ simulering för bayesiansk modell

Jag använde  $\begin{pmatrix} \theta_i \\ \bar{\alpha}_i \\ \bar{\beta}_i \end{pmatrix}$  och  $\begin{pmatrix} y\bar{\theta}_i \\ y\bar{\alpha}_i \end{pmatrix}$  från metropolis som min punktskattning för tricks och runs parame-

trar för att simulera 5000 LCQ:<br/>ar som vi gjorde i 2(d). Det ger oss följande frequency bar plot fig<br/>.  $4\,$ 

# (e) Teori om modellen

Jag ritade en acyclisk riktad graf för att representera modellen där

- $\bullet \ \Upsilon_{\mathbf{i}} = [{}^{y}A_{i}, {}^{y}B_{i}]^{T}$
- $\Theta_i = [\Theta_i, A_i, B_i]^T$
- $\mathbf{Y}_i = [Y_i^{(1)}, Y_i^{(2)}]^T$  betecknar is två runbetyg
- $\mathbf{X}_i = [X_i^{(1)}, X_i^{(2)}, X_i^{(3)}, X_i^{(4)}]^T$  betecknar is fyra trickbetyg

Nu ska vi svara på några frågor.

• Varför antog vi att  $\Upsilon_i \perp \Theta_i | \mathbf{X}_i, \mathbf{Y}_i \text{ i } 3(d)$ ?

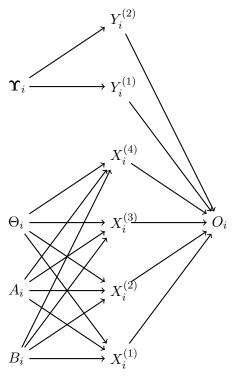


Figure 5: acyklisk riktad graf

• Kan vi dra slutsatsen från grafen fig. 5 att  $\Theta_i \perp A_i, B_i | \mathbf{X}_i$ ?

Nej. Betrakta stigen  $\Theta_i \to X_i^{(4)} \to A_i$ . För att  $X_i^{(4)} \in \operatorname{an}_G(\mathbf{X}_i)$  är  $\Theta$  och A, B d-sammanhängande.

• Kan vi dra slutsatsen från grafen fig. 5 att  $\Upsilon_i \perp \Theta_i | \mathbf{X}_i, \mathbf{Y}_i$ 

Ja för att  $\mathbf{X}_i, \mathbf{Y}_i \notin \mathrm{an}_G(O_i)$ 

• Kan vi anta fig. 5  $\Upsilon_i \perp \Theta_i | O_i$ ?

Nej för att  $O_i$  blockerar alla stigar mellan  $\Upsilon_i$  och  $\Theta_i$ . Intuitivt sett verkar det vettig att  $\Upsilon_i$  kan ge information om  $\Theta_i$  givet total betyget. Om vi vet lite om värdet av total betyg  $O_i$  och  $\Upsilon_i$  så vet vi ungefär vilka värde run betyg antar som kan ge information om trick betyget och därmed vilka värde  $\Theta_i$  kommer anta.

# 4 En bayesiansk modell med en hierarki

- (a) Apriori fördelning för  $[\Theta_i, C_i, D_i]^T$ 
  - Låt  $\Theta_i|C_i=c_i, D_i=d_i \sim \text{Beta}(c_i,d_i).$
  - Låt  $f_{c_i,d_i}(c_i,d_i) \propto (c_i+d_i)^{-5/2}$
- (b) Aposteriori för  $X_i$
- (c) Aposteriori för  $Y_i$
- (d) Simulering
- (e) Grafisk model

### 5 Diskussion

(a)

Hur jämför resultaten (skateboardåkarna i typvärdena) av de olika mod- ellerna? Vilka skateboardåkare är korrekt förutspådda och vilka inte är det? Ge några möjliga förklaringar till skillnaderna mellan de olika modellernas förutsägelser. Vilken modell föredrar du och varför?

(b)

Hur jämför dina skattningar för  $\theta_i$  i Uppgift 1 och dina skattade väntevär- den och varianser för  $\theta_i$  Uppgift 3 och 4? Vad är det förväntade betyget för ett trick för varje skateboardåkare givet att tricket har lyckats landa? Vad är det förväntade runbetyget? Med tanke på de skateboardåkare som förutspås vinna enligt de olika modellerna, ger dessa statistikor några insikter om framgångsrika strategier för att vinna? Till exempel, fungerar det att fokusera på ett bra runbetyg framför bra trickbetyg? Finns det exempel där denna strategi fungerar? Är det bra att ha bättre trickbetyg med stor varians eller lite sämre trickbetyg med mindre varians? osv.

(c)

Skatta väntevärdet och standardavvikelsen för varje skateboardåkares total- betyg för modellerna i uppgift 3 och 4. Stödjer denna statistik dina förut- sägelser? Enligt denna statistik, vad måste hända för att resultatet ska bli de riktiga vinnarna?

(d)

I alla modellerna antog vi att skateboardåkarens prestationer är oberoende. Till exempel antog vi att alla Vi är oberoende. Verkar detta som ett rimligt antagande? Motivera ditt svar.

(e)

I alla modeller struntade vi i ordningen som skateboardåkarna turades om. Verkar detta vara en rimlig sak att göra? Varför eller varför inte?