1. Três experimentos binomiais independentes são executados de maneira sucessiva. Em cada um deles, mede-se o número de respostas positivas que um sujeito fornece em certo número de tentativas. Seja $X_i \sim \text{Bin}(n_i, \theta_i)$ o número de sucessos no experimento i. Um estímulo é fornecido no experimento do meio de forma que a probabilidade de sucesso muda no experimento do meio e retorna para o nível inicial no terceiro experimento. Isto é, assume-se que $\theta_1 = \alpha$, $\theta_2 = \alpha + \beta$, e que $\theta_3 = \alpha$. Sabemos que uma v.a. $Y \sim \text{Bin}(n, \theta)$ tem função de probabilidade $\mathbb{P}(Y = k) = n!/(k!(n-k)!)\theta^k(1-\theta)^{n-k}$.

   - Supondo que $\beta = 0$ (isto é, que o estímulo não tem nenhum efeito), encontre o MLE de α
   - Obtenha o MLE de α e β no caso geral.

   Solução: Seja $L(\alpha, \beta)$ a verossimilhança de $\boldsymbol{\theta} = (\alpha, \beta)$. Pela independência, podemos escrever

$$\begin{aligned}
L(\alpha,\beta) &= \mathbb{P}(X_1 = x_1, X_2 = x_2, X_3 = x_3|\alpha,\beta) \\
&= \mathbb{P}(X_1 = x_1|\alpha,\beta)\mathbb{P}(X_2 = x_2|\alpha,\beta)\mathbb{P}(X_3 = x_3|\alpha,\beta) \\
&= \frac{n_1!}{(n_1-x_1)!x_1!}\alpha^{x_1}(1-\alpha)^{n_1-x_1}\frac{n_2!}{(n_2-x_2)!x_2!}(\alpha+\beta)^{x_2}(1-\alpha-\beta)^{n_2-x_2}\frac{n_3!}{(n_3-x_3)!x_3!}\alpha^{x_3}(1-\alpha)^{n_3-x_3} \\
&= K\alpha^{x_1+x_3}(1-\alpha)^{n_1+n_3-x_1-x_3}(\alpha+\beta)^{x_2}(1-\alpha-\beta)^{n_2-x_2}
\end{aligned}$$

   onde K é uma expressão que não envolve nem α, nem β.

   Tomando log, obtemos a log-verossimilhança:

$$\ell(\alpha,\beta) = \log(K) + (x_1+x_3)\log(\alpha) + (n_1+n_3-x_1-x_3)\log(1-\alpha) + x_2\log(\alpha+\beta) + (n_2-x_2)\log(1-\alpha-\beta)$$

   Considerando $\beta = 0$, temos uma log-verossimilhança apenas de α:

$$\begin{aligned}
\ell(\alpha,0) &= \log(K) + (x_1+x_3)\log(\alpha) + (n_1+n_3-x_1-x_3)\log(1-\alpha) + x_2\log(\alpha+0) + (n_2-x_2)\log(1-\alpha-0) \\
&= \log(K) + (x_1+x_3+x_2)\log(\alpha) + (n_1+n_3+n_2-x_1-x_3-x_2)\log(1-\alpha)
\end{aligned}$$

   Derivando em relação a α e igualando o resultado o zero temos:

$$0 = \frac{\partial\ell(\alpha,0)}{\partial\alpha} = \frac{x_1+x_3+x_2}{\alpha} + \frac{n_1+n_3+n_2-x_1-x_3-x_2}{1-\alpha}\ (-1)$$

   Isolando α do lado esquerdo encontramos o MLE de α *quando* $\beta = 0$: $\hat{\alpha}_0 = (x_1+x_3+x_2)/(n_1+n_3+n_2)$. Isto é, neste caso em que $\beta = 0$, o que temos éuma sequência de $n_1+n_2+n_3$ ensaios binários e independentes de Bernoulli em que a probabilidade de sucesso em cada ensaio é constante e igual a α. Neste caso, o MLE de α é simplesmente a proporção total de sucessos em todos os ensaios.

   Vamos agora considerar o caso geral em que β não é restrito a ser zero. Precisamos das derivadas parciais da log-verossimilhança $\ell(\alpha,\beta)$ com relação a cada um dos dois parâmetros. Derivando com relação a α e igualando a zero:

$$0 = \frac{\partial\ell(\alpha,\beta)}{\partial\alpha} = \frac{x_1+x_3}{\alpha} + \frac{n_1+n_3-x_1-x_3}{1-\alpha}\ (-1) + \frac{x_2}{\alpha+\beta} + \frac{n_2-x_2}{1-\alpha-\beta}\ (-1)$$

   Ou seja,

$$\frac{x_1+x_3}{\alpha} + \frac{x_2}{\alpha+\beta} = \frac{n_1+n_3-x_1-x_3}{1-\alpha} + \frac{n_2-x_2}{1-\alpha-\beta} \tag{1}$$

   Derivando com relação a β e igualando a zero:

$$0 = \frac{\partial\ell(\alpha,\beta)}{\partial\beta} \to \frac{x_2}{\alpha+\beta} = \frac{n_2-x_2}{1-\alpha-\beta} \to \beta = \frac{x_2}{n_2} - \alpha$$

   Substituindo esta expressão para β na primeira derivada (isto é, substituindo β por $x_2/n_2 - \alpha$ em (1)), encontramos um valor para o MLE de α:

$$\begin{aligned}
\frac{x_1+x_3}{\alpha} + \frac{x_2}{\alpha+x_2/n_2-\alpha} &= \frac{n_1+n_3-x_1-x_3}{1-\alpha} + \frac{n_2-x_2}{1-\alpha-(x_2/n_2-\alpha)} \\
\frac{x_1+x_3}{\alpha} + \frac{x_2}{x_2/n_2} &= \frac{n_1+n_3-x_1-x_3}{1-\alpha} + \frac{n_2-x_2}{1-x_2/n_2)} \\
\frac{x_1+x_3}{\alpha} + n_2 &= \frac{n_1+n_3-x_1-x_3}{1-\alpha} + n_2 \\
\frac{x_1+x_3}{\alpha} &= \frac{n_1+n_3-x_1-x_3}{1-\alpha} \\
\alpha &= \frac{x_1+x_3}{n_1+n_3}
\end{aligned}$$

Tendo o MLE de α voltamos na segunda derivada para encontrar o MLE de β:

$$\beta = \frac{x_2}{n_2} - \alpha = \frac{x_2}{n_2} - \frac{x_1 + x_3}{n_1 + n_3}$$

No final, encontramos que o MLE $\hat{\alpha}_1$ da chance de sucesso α no primeiro e terceiro experimentos é simplesmente a proporção $(x_1+x_3)/(n_1+n_3)$ de sucessos nestes dois experimentos. O parâmetro β mede o acréscimo nessa probabilidade de sucesso durante o segundo experimento devido ao estímulo aplicado. O MLE $\hat{\beta}$ desse parâmetro β é a diferença entre a proporção de sucessos neste segundo experimento e a proporção observada nos outros dois experimentos. Tudo isto é intuitivo e seria o que provavelmente você faria para estimar os parâmetros sem recorrer a teorias complicadas. A situação não será tão intuitivamente óbvia no próximo problema.

---

2. No problema anterior, assuma que o estímulo é aumentado no terceiro experimento e que $\theta_3 = \alpha + 2\beta$, além de termos $\theta_1 = \alpha$ e $\theta_2 = \alpha + \beta$. Monte a equação de iteração de Newton para atualizar os parâmetros obtendo o gradiente da log-verossimilhança $\ell(\alpha, \beta)$. Deixe APENAS indicada a dimensão da matriz Hessiana (não precisa calculá-la) mas esclarecendo como faria para obê-la. Supondo que o valor inicial seja $(\alpha^{(0)}, \beta^{(0)}) = (\bar{x}, 0)$ com $\bar{x} = (x_1 + x_2 + x_3)/(n_1 + n_2 + n_3)$, diga como faria para obter os dois primeiros passos do algoritmo.

Solução: A log-verossimilhança desta situação é obtida do memso modo que fizemos anteriormente mas fazendo $\theta_3 = \alpha + 2\beta$ no terceiro experimento:

$$\ell(\alpha,\beta) = \log(K) + x_1\log(\alpha) + (n_1-x_1)\log(1-\alpha) + x_2\log(\alpha+\beta) + (n_2-x_2)\log(1-\alpha-\beta) + x_3\log(\alpha+2\beta) + (n_3-x_3)\log(1-\alpha-2\beta)$$

O vetor gradiente é formado pelas duas derivadas parciais:

$$\nabla\ell(\alpha,\beta) = \left[\begin{array}{c} \frac{\partial\ell}{\partial\alpha} \\ \frac{\partial\ell}{\partial\beta} \end{array}\right] = \left[\begin{array}{c} \frac{x_1}{\alpha} - \frac{n_1-x_1}{1-\alpha} + \frac{x_2}{\alpha+\beta} - \frac{n_2-x_2}{1-\alpha-\beta} + \frac{x_3}{\alpha+2\beta} - \frac{n_3-x_3}{1-\alpha-2\beta} \\ \frac{x_2}{\alpha+\beta} - \frac{n_2-x_2}{1-\alpha-\beta} + \frac{2x_3}{\alpha+2\beta} - \frac{2(n_3-x_3)}{1-\alpha-2\beta} \end{array}\right]$$

Tomando bastante cuidado com os sinais ao derivar $1/(1-\alpha)$, a matriz Hessiana com as derivadas parciais de segunda ordem é a seguinte:

$$\begin{aligned}
\mathbf{H} &= \left[\begin{array}{cc} \frac{\partial^2\ell}{\partial\alpha^2} & \frac{\partial^2\ell}{\partial\alpha\beta} \\ \frac{\partial^2\ell}{\partial\beta\alpha} & \frac{\partial^2\ell}{\partial\beta^2} \end{array}\right] \\
&= -\left[\begin{array}{cc} \frac{x_1}{\alpha^2} + \frac{n_1-x_1}{(1-\alpha)^2} + \frac{x_2}{(\alpha+\beta)^2} + \frac{n_2-x_2}{(1-\alpha-\beta)^2} + \frac{x_3}{(\alpha+2\beta)^2} + \frac{n_3-x_3}{(1-\alpha-2\beta)^2} & \frac{x_2}{(\alpha+\beta)^2} + \frac{n_2-x_2}{(1-\alpha-\beta)^2} + \frac{2x_3}{(\alpha+2\beta)^2} + \frac{2(n_3-x_3)}{(1-\alpha-2\beta)^2} \\ \frac{x_2}{(\alpha+\beta)^2} + \frac{n_2-x_2}{(1-\alpha-\beta)^2} + \frac{2x_3}{(\alpha+2\beta)^2} + \frac{2(n_3-x_3)}{(1-\alpha-2\beta)^2} & \frac{x_2}{(\alpha+\beta)^2} + \frac{n_2-x_2}{(1-\alpha-\beta)^2} + \frac{4x_3}{(\alpha+2\beta)^2} + \frac{4(n_3-x_3)}{(1-\alpha-2\beta)^2} \end{array}\right]
\end{aligned}$$

A equação de iteração de Newton para encontrar o MLE $\hat{\boldsymbol{\theta}}$ do vetor $\boldsymbol{\theta} = (\alpha, \beta)$ é a seguinte:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \mathbf{H}^{-1}\ \nabla\ell(\alpha,\beta)$$

onde $\mathbf{H}$ e $\nabla\ell(\alpha, \beta)$ são avaliados no valor corrente $\boldsymbol{\theta}^{(t)}$. Um valor inicial poderia assumir um efeito de acréscimo nulo (isto é, $\beta^{(0)} = 0$) e $\alpha^{(0)} = (x_1 + x_2 + x_3)/(n_1 + n_2 + n_3)$, a média global.

---

3. A análise fatorial apresentada abaixo vem de um estudo de diferenças de linguagem em personagens de programas infantis de televisão. Os pesquisadores primeiramente pediram aos estudantes universitários que classificassem as transcrições de diálogos de personagens de TV em programas como "Castelo Rá-Tim-Bum" ou "Sítio do Pica-pau Amarelo" em 12 itens diferenciais semânticos. Cada item é uma variável assumindo um valor entre -10 e +10. Os itens/variáveis são:

- $V1$ = agradável / desagradável
- $V2$ = rico / pobre
- $V3$ = agressivo / não agressivo
- $V4$ = doce / azedo
- $V5$ = forte / fraco
- $V6$ = alto status social / baixo status social
- $V7$ = alfabetizado / analfabeto

- $V8$ = estridente / suave
- $V9$ = legal / horrível
- $V10$ = ativo / passivo
- $V11$ = trabalhador braçal / nã-braçal,
- $V12$ = bonito / feio

Em seguida, os pesquisadores analisaram esses dados com o modelo de análise fatorial. Usando os dados dos estudantes como vetores de dimensão 12, foi calculada a matriz de correlação 12×12 entre os itens para ver se eles poderiam ser representados por um conjunto menor de fatores latentes ou subjacentes. Consulte os dados abaixo para responder às seguintes perguntas.

Correlation Matrix

		VAR01	VAR02	VAR03	VAR04	VAR05	VAR06	VAR07	VAR08	VAR09
Correlation	VAR01	1.00000	.24205	-.02264	.57201	.25033	.23812	.44151	-.31878	.71302
	VAR02	.24205	1.00000	.17757	.26074	.26381	.76259	.44463	.01354	.20545
	VAR03	-.02264	.17757	1.00000	-.11690	.63075	.17730	.16794	.56848	-.15958
	VAR04	.57201	.26074	-.11690	1.00000	.06706	.25066	.30289	-.27833	.61851
	VAR05	.25033	.26381	.63075	.06706	1.00000	.29412	.33024	.35753	.14280
	VAR06	.23812	.76259	.17730	.25066	.29412	1.00000	.47306	-.00663	.22557
	VAR07	.44151	.44463	.16794	.30289	.33024	.47306	1.00000	-.13852	.41723
	VAR08	-.31878	.01354	.56848	-.27833	.35753	-.00663	-.13852	1.00000	-.40543
	VAR09	.71302	.20545	-.15958	.61851	.14280	.22557	.41723	-.40543	1.00000
	VAR10	.02765	.24435	.69403	.01502	.60833	.24400	.16177	.57160	-.05209
	VAR11	.24963	.59653	.14734	.18686	.24603	.62276	.48356	-.12353	.26732
	VAR12	.39584	.41994	.04516	.40189	.20847	.44424	.41434	-.09726	.32917

		VAR10	VAR11	VAR12
Correlation	VAR10	1.00000	.14009	.12410
	VAR11	. 14009	1.00000	.32406
	VAR12	.12410	. 32406	1.00000

	Component		
	1	2	3
VAR01	.13877	.86762	.02793
VAR02	.86337	.10739	.12294
VAR03	.11268	-.08149	.87381
VAR04	.14516	.77321	-.07247
VAR05	.19290	.24766	.78707
VAR06	.87911	.11943	.12498
VAR07	.55580	.45685	.14664
VAR08	-.08253	-.39076	.73684
VAR09	.13050	.87344	-.11956
VAR10	.14153	.00941	.86693
VAR11	.81605	.11813	.03101
VAR12	.47390	.46986	.07011

Variable Name:

VAR01 = pleasing/displeasing
VAR02 = rich/poor
VAR03 = aggressive/unaggressive
VAR04 = sweet/sour
VAR05 = strong/weak
VAR06 = hi social status/low social status
VAR07 = literate/illiterate
VAR08 = loud/soft
VAR09 = nice/awful
VAR10 = active/passive
VAR11 = white collar/blue collar
VAR12 = beautiful/ugly

Figura 1: Amostra de um vetor aleatório (X, Y).

- Na matriz de correlação, quais os pares de variáveis altamente correlacionadas? (listar alguns pares).
- Na matriz de factor loadings (cargas dos fatores), quais variáveis estão correlacionadas com quais fatores?
- Quantos fatores são indicados por esta análise?
- Dê um nome para cada fator.
- Existem variáveis com uma carga muito baixa em todos os fatores que justifique descartá-las?

Solução: Os pares de variáveis com correlação superior a 0.60 *em valor absoluto* são os seguintes: $(1, 9), (6, 2), (5, 3), (10, 3), (9, 4)$.

Considerando um limiar de carga mínima igual a 0.6 (em valor absoluto), as variáveis com maior carga no primeiro fator s ao: $V2, V6, V7$ e $V11$. São as variáveis rico, status, alfabetizado, braçal. Indicam uma diferenciação baseada na hierarquia social de renda e status. No segundo fator, as de maior carga são: $V1, V4$, e $V9$ (agradável, doce, legal), um fator que indica o quão agradável era o diálogo. No terceiro fator, as variáveis de maior carga são: $V3, V5, V8$ e $V10$ (agressivo, forte, estridente, ativo), um fator indicando intensidade ou força física ou psicológica.

Quantos fatores são indicados por esta análise? Não coloquei elementos para responder a esta pergunta. Erro meu. Não considerei as respostas a esta questão.

Nomes descritivos para os fatores: ver acima

Não, nenhuma variável possui cargas desprezíveis nos três fatores que justificaria descartá-la.