Avaliação suplementar (AS)

Disciplina: Modelagem Estatística
Instrutor: Luiz Max Carvalho
Monitores: Eduardo Adame & Ezequiel Braga

03 de julho de 2024

- O tempo para realização da prova é de 3 horas;
- Leia a prova toda com calma antes de começar a responder;
- Responda todas as questões sucintamente;
- Marque a resposta final claramente com um quadrado, círculo ou figura geométrica de sua preferência;
- A prova vale 100 pontos. A pontuação restante é contada como bônus;
- Apenas tente resolver a questão bônus quando tiver resolvido todo o resto;
- Você tem direito a trazer **uma folha de "cola"** tamanho A4 frente e verso (impressa ou escrita à mão), que deverá ser entregue junto com as respostas da prova.

1. There are more things in Heaven and Earth, Horatio...

Em muitos problemas estatísticos, temos erros de medição ou observação. Em Epidemiologia, por exemplo, raramente observamos o verdadeiro número de casos de uma doença; em vez disso, temos acesso aos números de casos **notificados**. Aqui vamos estudar alguns aspectos de modelos de contagem com erro de observação, em particular com variáveis *latentes*, que não são observáveis. Suponha que $X_1, X_2, \ldots, X_n$ são contagens ($X_i \in \mathbb{N} \cup \{0\}$), assumidas i.i.d. Suponha ainda que para $\mu, \phi > 0$, postulamos o seguinte modelo para os dados:

$$Y_i \sim \text{Binomial-Negativa}(\mu, \phi), \tag{1}$$
$$X_i \mid Y_i \sim \text{Binomial}(Y_i, \theta), \tag{2}$$

para $i = 1, 2, \ldots, n$, onde a binomial negativa é parametrizada em termos de média e dispersão. Lembre-se de que se $Y \sim \text{Binomial-Negativa}(\mu, \phi)$, temos

$$\Pr(Y = y \mid \mu, \phi) = \binom{y + \phi - 1}{y} \left(\frac{\phi}{\mu + \phi}\right)^{\phi} \left(\frac{\mu}{\mu + \phi}\right)^{y}, \tag{3}$$

$\mathbb{E}[W] = \mu$ e $\text{Var}(W) = \mu + \mu^2/\phi$.

a) (10 pontos) Suponha que observamos $S_n := \sum_{i=1}^{n} X_i$. Poderíamos, em tese, tentar uma estimação por método dos momentos. Para ajudar nesta tarefa, compute $\mathbb{E}[S_n]$ e $\text{Var}(S_n)$.

   Dica: Use a propriedade da torre e as leis de esperanças e variâncias.

b) (5 pontos) Calcule a probabilidade marginal conjunta $\Pr(X_1 = x_1, \ldots, X_n = x_n)$.

c) (10 pontos) O modelo em questão é identificável? Justifique sua resposta.

d) (10 pontos) Como você modificaria $\Pr(X_1 = x_1, \ldots, X_n = x_n)$ se o evento $X_i = 0$ não fosse observável?

e) (15 pontos) Suponha que o modelo fosse modificado da seguinte forma:

$$N \sim \text{Binomial-Negativa}(\mu, \phi),$$
$$X_i \mid N \sim \text{Binomial}(N, \theta).$$

   Como você responderia aos itens a, b e d?

Conceitos trabalhados: modelos de contagem, variáveis latentes, identificabilidade, truncamento. **Nível de dificuldade**: médio.

Resolução: Para o primeiro item, usando a propriedade da torre, observe que

$$\begin{aligned}
\mathbb{E}[X_i] &= \mathbb{E}[\mathbb{E}[X_i \mid Y_i]], \\
&= \mathbb{E}[Y_i \theta], \\
&= \mu\theta.
\end{aligned}$$

Agora, usando a linearidade do valor esperado, conclui-se que $\mathbb{E}[S_n] = n\mu\theta$. Para a variância, podemos usar a lei da variância total, ou seja,

$$\begin{aligned}
\operatorname{Var}(X_i) &= \operatorname{Var}(\mathbb{E}[X_i \mid Y_i]) + \mathbb{E}[\operatorname{Var}(X_i \mid Y_i)], \\
&= \operatorname{Var}(Y_i\theta) + \mathbb{E}[Y_i\theta(1-\theta)], \\
&= \mu\theta^2 + \frac{\mu^2\theta}{\phi} + \mu\theta(1-\theta), \\
&= \frac{\mu^2\theta}{\phi} + \mu\theta.
\end{aligned}$$

Agora, basta usar a suposição de iid e concluir que $\operatorname{Var}(S_n) = \frac{n\mu^2\theta}{\phi} + n\mu\theta$. Usando novamente essa premissa, podemos responder o item b observando que $\Pr(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^n \Pr(X_i = x_i)$ e que

$$\begin{aligned}
\Pr(X_i = x) &= \sum_{k=x}^{\infty} \Pr(Y_i = k) \Pr(X_i = x \mid Y_i = k), \\
&= \sum_{k=x}^{\infty} \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \binom{k+\phi-1}{k} \left(\frac{\mu}{\mu+\phi}\right)^{k} \binom{k}{x} \theta^x (1-\theta)^{k-x}, \\
&= \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\theta}{1-\theta}\right)^{x} \sum_{k=x}^{\infty} \binom{k}{x} \binom{k+\phi-1}{k} \left(\frac{\mu(1-\theta)}{\mu+\phi}\right)^{k}, \\
&= \binom{x+\phi-1}{x} \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\theta}{1-\theta}\right)^{x} \left(\frac{\mu(1-\theta)}{\mu+\phi}\right)^{x} \left(\frac{\theta\mu+\phi}{\mu+\phi}\right)^{-(\phi+x)}, \\
&= \binom{x+\phi-1}{x} \left(\frac{\phi}{\theta\mu+\phi}\right)^{\phi} \left(\frac{\theta\mu}{\theta\mu+\phi}\right)^{x}.
\end{aligned}$$

Logo, $X_i \sim$ Binomial-Negativa$(\theta\mu, \phi)$ e

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \binom{x_i+\phi-1}{x_i} \left(\frac{\phi}{\theta\mu+\phi}\right)^{\phi} \left(\frac{\theta\mu}{\theta\mu+\phi}\right)^{x_i}.$$

Para a identificabilidade, note que se tomarmos $\theta_1 = \frac{1}{\mu_1}$ e $\theta_2 = \frac{1}{\mu_2}$, com $\mu_1 \neq \mu_2$, temos $\Pr_{\theta_1,\mu_1,\phi} = \Pr_{\theta_2,\mu_2,\phi}$, mas $\theta_1 \neq \theta_2$ e $\mu_1 \neq \mu_2$, ou seja, o modelo não é identificável.

Se $X_i = 0$ não for observável, podemos fazer um truncamento, ou seja, definimos uma variável Z_i tal que

$$\Pr(Z_i = z) = \frac{\Pr(X_i = z)}{1 - \Pr(X_i = 0)}, \; z > 0.$$

A partir disso, as contas anteriores podem ser refeitas.

Para o último item, observe que agora as variáveis deixam de ser iid, mas são condicionalmente independentes. Note que $\mathbb{E}[S_n]$ permanece o mesmo e, usando a lei da variância total,

$$\begin{aligned}\operatorname{Var}(S_n) &= \operatorname{Var}(nN\theta) + \mathbb{E}[nN\theta(1-\theta)],\\ &= (n\theta)^2(\mu + \mu^2/\phi) + n\mu\theta(1-\theta).\end{aligned}$$

Para a probabilidade conjunta, temos

$$\begin{aligned}\Pr(X_1 = x_1, \ldots, X_n = x_n \mid N = k) &= \prod_{i=1}^{n} \Pr(X_i = x_i \mid N = k),\\ &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{nk - \sum_{i=1}^n x_i} \prod_{i=1}^{n} \binom{k}{x_i}.\end{aligned}$$

Então,

$$\begin{aligned}\Pr(X_1 = x_1, \ldots, X_n = x_n) &= \sum_{k=x_{(n)}}^{\infty} \Pr(X_1 = x_1, \ldots, X_n = x_n \mid N = k) \Pr(N = k),\\ &= \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n x_i} \sum_{k=x_{(n)}}^{\infty} \binom{k+\phi-1}{k} \left(\frac{\mu(1-\theta)^n}{\mu+\phi}\right)^{k} \prod_{i=1}^{n} \binom{k}{x_i}.\end{aligned}$$

Para realizar o truncamento, podemos considerar $Z_i \mid N$ tal que

$$\Pr(Z_i = z \mid N) = \frac{\Pr(X_i = z \mid N)}{1 - \Pr(X_i = 0 \mid N)}, z > 0.$$

■

Comentário: Nessa questão vimos como usar princípios básicos de probabilidade para construir modelos probabilísticos. Vimos como adequar o modelo observacional aos dados, marginalizando para expressar a verossimilhança apenas em função dos dados que de fato são observados. Vimos também como acomodar truncamento, que é um outro problema clássico de erro de observação. Finalmente, vimos como modificar o modelo alterando sua estrutura latente (de n variáveis latentes para apenas uma) e as implicações que isso tem.

2. Will Zelda like this one?

Os modelos lineares generalizados são um poderoso cavalo de batalha da Estatística Aplicada. Nesta questão vamos explorar a conexão entre modelos de contagem e modelos de regressão binária. Tome $\boldsymbol{X}$ uma matriz $n \times P$ de covariáveis e suponha que $Z_i \sim \text{Poisson}(\mu_i)$ para $i = 1, 2, \ldots, n$ são amostras independentes. Suponha ainda que temos um GLM Poisson para modelar $E[Z_i]$ usando a função de ligação canônica. Denote os coeficientes deste modelo de $\boldsymbol{\beta}$.

a) (10 pontos) Suponha que estamos interessados em calcular a mudança percentual na média quando a covariável X_j muda uma unidade, mantidas todas as outras constantes. Exiba um estimador dessa quantidade a partir de um estimador $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_P)$ dos coeficientes.

b) (20 pontos) Ainda considerando o modelo anterior, defina $Y_i = \mathbb{I}(Z_i > 0)$ e defina $\theta_i = \Pr(Y_i = 1)$.

   (a) Mostre que $Y_i \sim \text{Bernoulli}(\theta_i)$;

   (b) Exiba a g tal que $g(\theta_i) = \boldsymbol{X}_i\boldsymbol{\beta}$.

Conceitos trabalhados: regressão binária, GLM, função de ligação. **Nível de dificuldade**: médio.

Resolução: É fácil ver que a função de ligação canônica é log neste caso. Logo, a mudança em uma unidade em relação a $\log(\mu_i)$ na covariável X_j (mantendo todas as outras constantes) é β_j. Então, um estimador da mudança esperada em μ_j é $\exp(\hat{\beta}_j)$.

Para encontrar a distribuição de Y_i, note que $Y_i \in \{0, 1\}$ e que

$$\Pr(Y_i = 1) = \Pr(Z_i > 0),$$
$$\Pr(Y_i = 0) = \Pr(Z_i = 0),$$

o que conclui o desejado. Por outro lado, se $\log \mu_i = \boldsymbol{X_i\beta}$ e $\theta_i = \Pr(Y_i = 1) = 1 - \exp(-\mu_i)$, temos $\mu_i = -\log(1 - \theta_i)$ e, então, $\log[-\log(1 - \theta_i)] = \boldsymbol{X}_i\boldsymbol{\beta}$. ■

Comentário: Nesta questão vimos a conexão entre GLM Poisson e uma família um pouco menos conhecida de modelos com a função de ligação log-log complementar, também chamada de Gompertz. O interessante é que esse modelo pode então ser usado para fazer cálculos sobre a variável indicadora ou as contagens, a depender dos dados que estavam disponíveis para ajustar o modelo. Isso é útil, por exemplo, na modelagem de dados de sobrevivência.

3. There's levels to this game

Modelos multinível são de grande importância em modelagem estatística, possibilitando a representação de problemas onde as observações se dividem em grupos. Suponha que m escolas são escolhidas entre milhares de cidades no Brasil. Dentro de cada escola, são escolhidos n estudantes da mesma idade são selecionados e suas notas em um teste padronizado são anotadas. Seja Y_{ij} a nota do aluno j na escola i. Um modelo simples é

$$Y_{ij} = \mu + U_i + \varepsilon_{ij},$$

onde μ é o valor esperado da nota na população geral, U_i é um *efeito aleatório* por escola e ε_{ij} é um termo de erro ao nível do indivíduo. Vamos suplementar esse modelo fazendo as seguintes premissas adicionais:

$$U_i \sim \text{Normal}(0, \sigma^2),$$
$$\varepsilon_{ij} \sim \text{Normal}(0, \tau^2),$$

onde $\sigma^2, \tau^2 > 0$ são as variâncias do efeito aleatório e dos resíduos, respectivamente.

Em particular,

a) (10 pontos) Suponha que queremos estimar μ, σ^2 e τ^2 por máxima verossimilhança. Isso é complicado por causa da presença dos efeitos aleatórios $\boldsymbol{U} = (U_1, \ldots, U_m)$, que não são observados[1]. Uma alternativa é utilizar a **verossimilhança marginal**. Encontre a densidade marginal condicional conjunta de $\boldsymbol{Y} = (Y_{11}, \ldots, Y_{n1}, Y_{12} \ldots, Y_{nm})$, $f(\boldsymbol{Y} \mid \mu, \sigma^2, \tau^2)$.

b) (10 pontos) Na *análise de variância (ANOVA)*, o interesse é em estimar componentes da variância intra-grupo e entre grupos. Defina a média de cada escola como $\bar{Y}_i := \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ e a grande média como $\bar{\bar{Y}} := \frac{1}{m}\sum_{i=1}^{m} \bar{Y}_i$. Com essas quantidades podemos computar as somas de quadrados **dentro** de cada grupo,

$$W := \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_i\right)^2,$$

e **entre** grupos,

$$B := n\sum_{i=1}^{m}\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2.$$

Exiba um estimador não-viesado de τ^2 baseado em W.

c) * (20 pontos extra) Sugira um estimador para a correlação intraclasse

$$\text{ICC} := \frac{\sigma^2}{\sigma^2 + \tau^2},$$

baseado em W e B. O estimador proposto pode ser viesado.

Conceitos trabalhados: modelos multinível, componentes de variância, correlação intra-classe. **Nível de dificuldade**: difícil.

Resolução: Como as observações são assumidas independentes, podemos escrever a verossimilhança marginal como

$$f(\boldsymbol{Y} \mid \mu, \sigma^2, \tau) = \prod_{i=1}^{m}\prod_{j=1}^{n}\int h_{Y_{ij}}(y_{ij}; \mu, \sigma, \tau, u_i) g_{U_i}(u_i; \sigma) du_i,$$

onde h é pdf condicional de $Y_{ij} \mid U_i$, que é uma normal com média $\mu + U_i$ e variância τ^2; e g, a pdf de U_i, também normal, com média 0 e variância τ^2. De maneira mais direta, basta observar que Y_{ij} é a soma de normais independentes e, portanto, a integral acima resulta na pdf de uma normal de média μ e variância $\sigma^2 + \tau^2$.

Para o item b, note que $\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2 = \sum_{j=1}^{n}(Y_{ij} - \mu)^2 - n(\bar{Y}_i - \mu)^2$. Tomando o valor esperado, temos

[1] Nem observáveis.

$$
\begin{aligned}
\mathbb{E}\left[\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_i)^2\right] &= \sum_{j=1}^{n}\mathrm{Var}(Y_{ij}) - n\,\mathrm{Var}(\bar{Y_{ij}}),\\
&= n(\sigma^2+\tau^2) - \frac{1}{n}\left(\sum_{j=1}^{n}\mathrm{Var}(Y_{ij}) + \sum_{j\neq k}\mathrm{Cov}(Y_{ij},Y_{ik})\right),\\
&= n(\sigma^2+\tau^2) - \frac{1}{n}\left(n(\sigma^2+\tau^2)+n(n-1)\sigma^2\right),\\
&= (n-1)(\sigma^2+\tau^2) - (n-1)\sigma^2,\\
&= (n-1)\tau^2.
\end{aligned}
$$

Observe que na terceira linha usamos que, para $j \neq k$,

$$
\begin{aligned}
\mathrm{Cov}(Y_{ij},Y_{ik}) &= \mathrm{Cov}(\mu+U_i+\varepsilon_{ij},\mu+U_i+\varepsilon_{ik}),\\
&= \mathrm{Var}(U_i),\\
&= \sigma^2.
\end{aligned}
$$

Logo, $\mathbb{E}[W] = m(n-1)\tau^2$ e, portanto, $\widehat{\tau^2} = \frac{1}{m(n-1)}W$ é um estimador não-viesado de τ^2.

Para o último item, vamos encontrar um estimador não-viesado de σ^2 baseado em em B e W. Assim, observe que

$$
\mathbb{E}\left[\bar{Y}_i - \bar{\bar{Y}}\right] = \mathbb{E}\left[\bar{Y}_i\right] - \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\left[\bar{Y}_i\right],
$$

e que

$$
\begin{aligned}
\mathbb{E}\left[\bar{Y}_i\right] &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[Y_{ij}\right],\\
&= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\mu+U_i+\varepsilon_{ij}\right],\\
&= \mu.
\end{aligned}
$$

Logo,

$$
\begin{aligned}
\mathbb{E}[B] &= n\sum_{i=1}^{m}\mathrm{Var}\left(\bar{Y}_i-\bar{\bar{Y}}\right)\\
&= n\sum_{i=1}^{m}\left[\mathrm{Var}\left(\bar{Y}_i\right)+\mathrm{Var}\left(\bar{\bar{Y}}\right)-2\mathrm{Cov}\left(\bar{Y}_i,\bar{\bar{Y}}\right)\right].
\end{aligned}
$$

Agora, note que

$$
\begin{aligned}
\operatorname{Var}\left(\bar{Y}_i\right) &= \operatorname{Var}\left(\mu + U_i + \frac{1}{n}\sum_{j=1}^{n}\varepsilon_{ij}\right), \\
&= \operatorname{Var}\left(U_i\right) + \frac{1}{n}\operatorname{Var}\left(\varepsilon_{i1}\right), \\
&= \sigma^2 + \frac{\tau^2}{n}.
\end{aligned}
$$

Além disso,

$$
\begin{aligned}
\operatorname{Var}\left(\bar{\bar{Y}}\right) &= \frac{1}{m^2}\operatorname{Var}\left(\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}(\mu + U_i + \varepsilon_{ij})\right), \\
&= \frac{1}{m^2}\operatorname{Var}\left(m\mu + \sum_{i=1}^{m}U_i + \sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\varepsilon_{ij})\right), \\
&= \frac{\sigma^2}{m} + \frac{1}{m^2n^2}\operatorname{Var}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\varepsilon_{ij})\right), \\
&= \frac{\sigma^2}{m} + \frac{\tau^2}{mn}.
\end{aligned}
$$

Por fim,

$$
\begin{aligned}
\operatorname{Cov}\left(\bar{Y}_i, \bar{\bar{Y}}\right) &= \frac{1}{m}\sum_{k=1}^{m}\operatorname{Cov}\left(\bar{Y}_i, \bar{Y}_k\right), \\
&= \frac{1}{m}\sum_{k=1}^{m}\operatorname{Cov}\left(\bar{Y}_i, \bar{Y}_i\right), \\
&= \frac{1}{m}\left(\sigma^2 + \frac{\tau^2}{n}\right).
\end{aligned}
$$

Combinando estes resultados, concluímos que $\mathbb{E}[B] = (m-1)(n\sigma^2 + \tau^2)$. Então, um estimador não-viesado para σ^2 pode ser obtido como

$$
\begin{aligned}
\widehat{\sigma^2} &= \frac{B}{n(m-1)} - \frac{\widehat{\tau^2}}{n}, \\
&= \frac{B}{n(m-1)} - \frac{W}{nm(n-1)}.
\end{aligned}
$$

Portanto, um estimador para ICC pode ser dado como

$$
\widehat{ICC} = \frac{\widehat{\sigma^2}}{\widehat{\sigma^2} + \widehat{\tau^2}}.
$$

■

Comentário:. Os modelos hierárquicos formam um grande corpo de técnicas para o tratamento de dados estratificados. Nesta questão, analisamos um modelo hierárquico normal, onde o **efeito aleatório** também é normal. Como vimos, essa premissa permite a **marginalização** do efeito aleatório, que por sua vez permite a escrita da verossimilhança apenas em função dos dados $\boldsymbol{Y}$ das notas dos alunos. Mais que isso, trabalhamos também as propriedades das somas de quadrados *intra*- e *inter*-grupos, e vimos como estimar uma estatístca (o ICC) que mede a força de associação do fator em questão (escolas) com a variável-resposta (notas) ao comparar a fração da variância que é explicada pelo fator analisado. *Absolute cinema*!