

Caso Gamma

Relatório

Silvaneio Viera dos Santos Junior

2022-10-31

Introdução

Neste relatório apresentaremos os resultados das análises feitas sobre o GDLM k-paramétrico para o caso Gamma com parâmetro de forma e média desconhecidos. Para esta análise, vamos assumir o seguinte modelo observacional:

$$X|\phi, \mu \sim \mathcal{G}\left(\phi, \frac{\phi}{\mu}\right),$$

onde \mathcal{G} representa a distribuição Gamma. Por conveniência, usaremos a parametrização com $\alpha = \phi$ e $\beta = \frac{\phi}{\mu}$ de modo que:

$$X|\alpha, \beta \sim \mathcal{G}(\alpha, \beta),$$

sendo que transitar de uma parametrização para a outra é trivial e a resolução dos sistemas de compatibilização é a mesma.

Para os parâmetros α e β , temos que a priori conjugada é tal que:

$$\pi(\alpha, \beta) \propto \exp\{n_0 \alpha \ln(\beta) - k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha - \tau_0 \beta\},$$

onde n_0, k_0, θ_0 e τ_0 são os parâmetros da distribuição e Γ é a função *Gamma*. Quando um par de variáveis aleatórias X, Y tiver a densidade descrita acima, diremos que $X, Y \sim \Pi(n_0, k_0, \theta_0, \tau_0)$, sendo que $n_0, k_0, \tau_0 > 0$. No caso especial onde $n_0 = k_0$, diremos que $X, Y \sim \Pi(n_0, \theta_0, \tau_0)$.

Ao obter uma amostra de tamanho m do modelo observacional, a obtenção dos parâmetros da posteriori $(n_m, k_m, \theta_m$ e $\tau_m)$ pode ser feita a partir das equações a seguir:

$$\begin{aligned} n_m &= n_0 + m \\ k_m &= k_0 + m \\ \theta_m &= \theta_0 + \sum_{i=1}^m \ln(x_i) \\ \tau_m &= \tau_0 + \sum_{i=1}^m x_i. \end{aligned}$$

Por último, esta distribuição pertence à família exponencial e o vetor de estatísticas suficientes associado a esta distribuição é:

$$H_p = (\alpha, \beta, \alpha \ln(\beta), \ln(\Gamma(\alpha)))'$$

Para utilizar o método proposto no artigo k-paramétrico é necessário obter $\mathbb{E}_p[H_p]$ e $\mathbb{E}_q[H_p]$, onde E_p é o valor esperado calculado com α e β tendo a distribuição conjugada p e E_q é o valor esperado calculado com α e β tendo distribuição log-Normal.

Na próxima sessão discutiremos algumas propriedades da distribuição Π , pois diversos problemas encontrados tem sua origem nas características de Π .

Na sessão subsequente abordaremos os resultados da tentativa de se calcular $\mathbb{E}_p[H_p]$ usando aproximações de Laplace (Tierney e Kadane, 1995). Infelizmente, não conseguimos obter um ajuste funcional com esta abordagem devido a problemas na solução do sistema $\mathbb{E}_p[H_p] = \mathbb{E}_q[H_p]$.

Na última sessão apresentamos uma proposta que permite obter uma expressão analítica aproximada para $\mathbb{E}_p[H_p]$. Com isso, conseguimos resolver o sistema (ainda usando Newton-Raphson, mas sem problemas numéricos) e fazer o ajuste do modelo. Ainda assim, o ajuste deixa a desejar. Mais investigações estão sendo feitas para tentar identificar o problema.

Propriedades da distribuição Π

Primeiro, observemos que, se $\alpha, \beta \sim \Pi(n_0, k_0, \theta_0, \tau_0)$, então:

$$\begin{aligned} f(\beta|\alpha) &\propto \pi(\alpha, \beta) \propto \exp \{n_0 \alpha \ln(\beta) - k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha - \tau_0 \beta\} \\ &\propto \exp \{n_0 \alpha \ln(\beta) - \tau_0 \beta\} = \beta^{n_0 \alpha} e^{-\tau_0 \beta}, \end{aligned}$$

ou seja $\beta|\alpha \sim \mathcal{G}(n_0 \alpha + 1, \tau_0)$.

Usando a distribuição condicional de β podemos reescrever $\mathbb{E}_p[H_p] = \mathbb{E}_p[\mathbb{E}_p[H_p|\alpha]]$, de onde obtemos:

$$\begin{aligned} \mathbb{E}_p[\beta] &= \mathbb{E}_p[\mathbb{E}_p[\beta|\alpha]] = \mathbb{E}_p \left[\frac{n_0 \alpha + 1}{\tau_0} \right] = \frac{n_0 \mathbb{E}_p[\alpha] + 1}{\tau_0} \\ \mathbb{E}_p[\alpha \ln(\beta)] &= \mathbb{E}_p[\alpha \mathbb{E}_p[\ln(\beta)|\alpha]] = \mathbb{E}_p[\alpha(\psi(n_0 \alpha + 1) - \ln(\tau_0))] \\ &= \mathbb{E}_p[\alpha \psi(n_0 \alpha + 1)] - \ln(\tau_0) \mathbb{E}_p[\alpha]. \end{aligned}$$

Usando que $\mathbb{E}_p[\alpha] = \mathbb{E}_q[\alpha]$ ($\mathbb{E}_q[H_p]$ é suposto conhecido), temos que:

$$n_0 = \frac{\mathbb{E}_q[\beta] \tau_0 - 1}{\mathbb{E}_q[\alpha]}$$

Com as equações acima, conseguimos escrever $\mathbb{E}_p[H_p]$ como valores esperados que dependem apenas da distribuição marginal de α , o que pode ser útil para simplificar algumas integrais e possibilitar a resolução numérica com métodos determinísticos. Vale observar que a distribuição marginal de α é tal que:

$$\begin{aligned}
f(\alpha) &\propto \int_0^{+\infty} \pi(\alpha, \beta) d\beta \\
&\propto \int_0^{+\infty} \exp \{n_0 \alpha \ln(\beta) - k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha - \tau_0 \beta\} d\beta \\
&= \exp \{-k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha\} \int_0^{+\infty} \exp \{n_0 \alpha \ln(\beta) - \tau_0 \beta\} d\beta \\
&= \exp \{-k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha\} \int_0^{+\infty} \beta^{n_0 \alpha + 1 - 1} e^{-\tau_0 \beta} d\beta \\
&= \exp \{-k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha\} \frac{\Gamma(n_0 \alpha + 1)}{\tau_0^{n_0 \alpha + 1}} \\
&= \exp \{-k_0 \ln(\Gamma(\alpha)) + \theta_0 \alpha + \ln(\Gamma(n_0 \alpha + 1)) - (n_0 \alpha + 1) \ln(\tau_0)\} \\
&\propto \exp \{\ln(\Gamma(n_0 \alpha + 1)) - k_0 \ln(\Gamma(\alpha)) + (\theta_0 - n_0 \ln(\tau_0)) \alpha\} \\
&= \frac{\Gamma(n_0 \alpha + 1)}{\Gamma(\alpha)^{k_0}} \exp \{(\theta_0 - n_0 \ln(\tau_0)) \alpha\}.
\end{aligned}$$

Usando a densidade acima e aproveitando a escrita de $\mathbb{E}_p[H_p]$ como uma valor esperado em α , podemos obter $\mathbb{E}_p[H_p]$ usando quadratura Gaussiana e dispensando o uso da aproximação de Laplace.

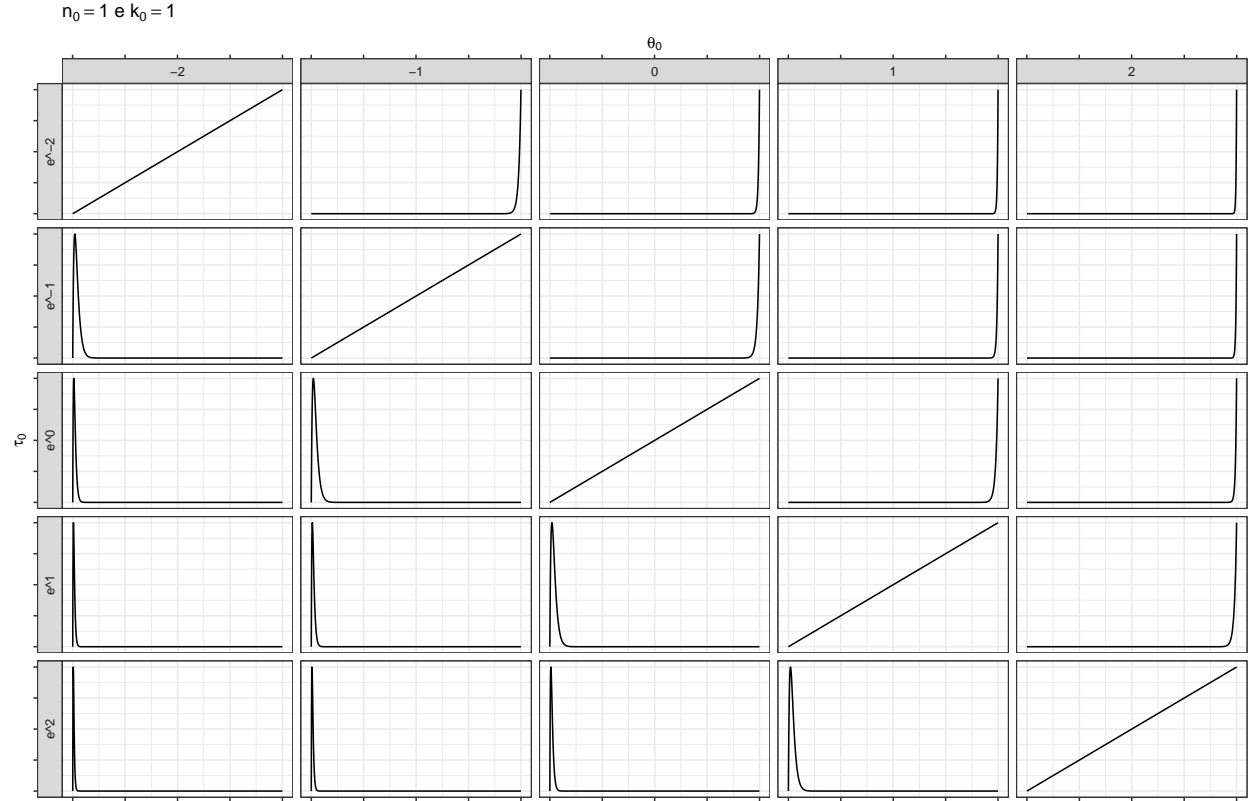
Como discutido em outras reuniões, um pre-requisito para o uso do Teorema da Projeção é que o valor esperado de H_p exista e seja finito. Caso esta condição não seja satisfeita, não podemos fazer a compatibilização das prioris normal e conjugada. Assim, devemos encontrar as condições para as quais Π é própria (i.e., a constante de normalização de π é finita) e o valor esperado de H_p é finito. Para facilitar esta análise, podemos fazer o estudo da distribuição marginal de α , pois se Π é própria, α também será, ademais, podemos avaliar o valor esperado de H_p olhando apenas para a distribuição de α .

Adiante, vamos exibir a densidade não normalizada de α para vários valores de k_0 , n_0 , θ_0 e τ_0 , porém, como são muitas combinações de parâmetros, a análise acaba se tornando exaustiva, por isso, antes de apresentar os gráficos, vamos resumir as conclusões:

- Se n_0 e k_0 são grandes em comparação a τ_0 e θ_0 a densidade de α se torna crescente em α a partir de algum valor de α (como consequência, a distribuição marginal de α não é própria). Observe que, para uma amostra grande, temos que $n_m \approx k_m \approx m$, $\tau_m \approx n_m \sum_{i=1}^m x_i / m$, $\theta_m \approx k_m \sum_{i=1}^m \ln(x_i) / m$, assim, a observação feita neste item equivale a dizer que a média dos x_i 's e dos $\ln(x_i)$'s não pode ser demasiadamente pequena.
- Se $n_0 \geq k_0$ a densidade de α se torna crescente em α para grande parte dos possíveis valores de τ_0 e θ_0 .

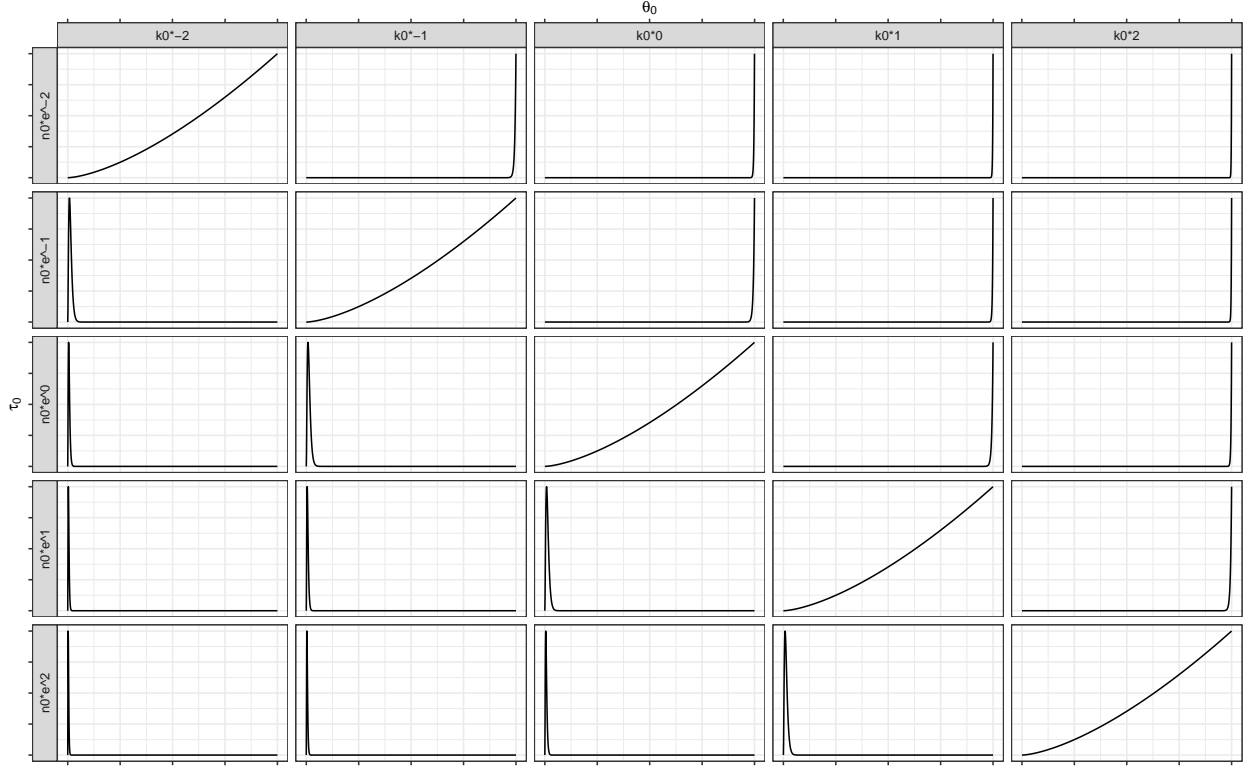
Para garantir que a distribuição marginal de α seja própria, precisamos que n_0 seja significativamente maior que k_0 (o quão menor vai depender da escala de k_0) e/ou que τ_0 e θ_0 estejam compatíveis com a escala de k_0 e n_0 . Em geral, essa informação será relevante apenas para inicialização do Newton-Raphson, sendo necessária uma escolha que evite que o algoritmo passe por “regiões ruins”.

A seguir, apresentamos a densidade marginal não normalizada de α para $n_0 = k_0 = 1$ e diversos valores de τ_0 e θ_0 :



Veja que, se $\tau_0 = e^{\theta_0}$, a densidade de α é simplesmente uma reta crescente em α . Se $\tau_0 > e^{\theta_0}$ a densidade de α é própria e se $\tau_0 \leq e^{\theta_0}$ a densidade de α é crescente em α (a partir de algum valor), logo a distribuição de α não é própria.

$$n_0 = 2 \text{ e } k_0 = 2$$



Com $n_0 = k_0 = 2$ temos um resultado parecido com o anterior, porém, o “ponto de corte” para tornar a densidade imprópria muda. De modo geral, quando $n_0 = k_0 = m$, observamos que o ponto de corte é $\ln\left(\frac{\tau_0}{m}\right) > \frac{\theta_0}{m}$. Intuitivamente, podemos entender a razão para este ponto de corte da seguinte forma: Se temos uma amostra de tamanho m com m muito grande, então $\tau_m \approx m \sum \frac{x_i}{m}$ e $\theta_m \approx m \sum \frac{\ln(x_i)}{m}$, então teríamos que $\ln\left(\frac{\tau_m}{m}\right) > \frac{\theta_m}{m}$, pois a função logarítmo é côncava, portanto $\ln\left(\frac{\tau_m}{m}\right) \approx \ln\left(\sum \frac{x_i}{m}\right) > \sum \frac{\ln(x_i)}{m} \approx \frac{\theta_m}{m}$. Ou seja, é “artificial” para uma Π que a condição $\ln\left(\frac{\tau_0}{m}\right) > \frac{\theta_0}{m}$ não seja válida, pois dados reais nunca produziram parâmetros sem essa propriedade.

A partir das análises feitas não conseguimos encontrar uma regra que garanta que a densidade de α seja própria a menos que tomemos $n_0 = k_0$, porém, ainda nesse caso, a restrição encontrada é inconveniente, pois a restrição $\ln\left(\frac{\tau_0}{m}\right) > \frac{\theta_0}{m}$ induz um espaço paramétrico onde não há garantias de que exista um elemento que minimize a divergência KL. No geral, tivemos muitos problemas em encontrar o mínimo, pois o algoritmo frequentemente sai do conjunto válido de parâmetros. Em diversas ocasiões conseguimos encontrar um valor inicial para os parâmetros de modo que o algoritmo de Newton-Raphson convirja de forma adequada, mas não conseguimos estabelecer um critério geral que garanta que sempre poderemos resolver o sistema. Uma forma de mitigar esse problema seria através da simplificação dos sistemas (se possível), pois isso facilitaria a busca dos parâmetros.

Método de Laplace

Suponhamos que queremos calcular $\mathbb{E}[g(x)]$ com x tendo densidade proporcional a f^* e g sendo uma função positiva, então podemos escrever:

$$\mathbb{E}[g(x)] = \frac{\int_{\mathbb{R}} g(x) f^*(x) dx}{\int_{\mathbb{R}} f^*(x) dx}.$$

Se considerarmos que:

$$g(x)f^*(x) \approx \exp \left\{ L_1(x_1^*) - \frac{(x - x_1^*)^2}{2v_1} \right\},$$

$$f^*(x) \approx \exp \left\{ L_2(x_2^*) - \frac{(x - x_2^*)^2}{2v_2} \right\},$$

onde $L_1(x) = \ln(g(x)f^*(x))$, $L_2(x) = \ln(f^*(x))$, x_i^* é o argumento que maximiza L_i e $v_i = -L_i''^{-1}(x_i^*)$. Usando a aproximação acima, obtemos:

$$\mathbb{E}[g(x)] \approx \left(\frac{v_1}{v_2} \right)^{\frac{1}{2}} \exp \{ L_1(x_1^*) - L_2(x_2^*) \}$$

A abordagem acima pode ser facilmente generalizada para o caso onde x é um vetor.

A ideia proposta pelo Migon é utilizar o método descrito para calcular $\mathbb{E}_p[H_p]$, porém encontramos um problema ao tentar por em prática esta proposta: Para calcular $\mathbb{E}_p[H_p]$ pelo método de Laplace precisamos conhecer os parâmetros da distribuição conjugada, mas desejamos calcular $\mathbb{E}[g(x)]$ justamente para encontrar os parâmetros da distribuição conjugada.

Para apresentar o problema de forma clara, vamos descrever o algoritmo de Newton-Raphson para a solução do sistema $\mathbb{E}_p[H_p] = \mathbb{E}_q[H_p]$:

- Passo 0: Suponha que conhecemos $\mathbb{E}_q[H_p]$ e seja $\phi = \mathbb{E}_p[H_p]$ com $p = \Pi(n, k, \tau, \theta)$.
- Passo 1: Inicializamos escolhendo n, k, τ e θ como valores válidos.
- Passo 2: Calculamos ϕ e $\nabla\phi$ (a matriz de derivadas parciais de ϕ com relação aos parâmetro de Π). Caso não exista a forma analítica para $\nabla\phi$, devemos avaliar ϕ 4 vezes (além da avaliação inicial) para calcular numericamente as derivadas de ϕ .
- Passo 3: Atualizamos n, k, τ e θ segundo o algoritmo de Newton-Raphson.
- Passo 4: Se ϕ é suficientemente próximo de $\mathbb{E}_q[H_p]$ encerramos o algoritmo, do contrário voltamos ao passo 2.

Usando o método de Laplace para obter ϕ , digamos, para encontrar $\mathbb{E}_p[\alpha]$, então temos que $g(\alpha, \beta) = \alpha$ e devemos encontrar α_1^* , β_1^* , α_2^* e β_2^* que maximizam L_1 e L_2 , respectivamente. Infelizmente, não é possível obter uma forma analítica fechada para α_1^* e β_1^* , de modo que seria necessário usar o método de Newton-Raphson para encontrar esses valores, porém isto está ocorrendo dentro de **uma** avaliação de **uma** das componentes de ϕ para **uma** iteração do método de Newton-Raphson, ou seja, seria necessário usar o método de Newton-Raphson 20 vezes **para cada iteração** do método de Newton-Raphson principal (temos de calcular ϕ 5 vezes a cada iteração e em cada cálculo precisamos usar Newton-Raphson 4 vezes). Por conta disso, é inviável usar a abordagem acima para realizar a compatibilização das prioris. Se fosse possível obter α_1^* , β_1^* de forma analítica para todos os parâmetro, não haveria problema, porém, não sendo este o caso, se torna inviável a resolução do sistema.

Como alternativa ao método descrito acima, podemos usar o seguinte fato, se $\Pi(n, k, \theta, \tau)$ pertence à família exponencial, então:

$$\mathbb{E}_p[H_p] = \nabla A(n, k, \theta, \tau),$$

onde $A(n, k, \theta, \tau)$ é o logaritmo da constante de normalização de $\Pi(n, k, \theta, \tau)$, isto é:

$$\exp\{A(n, k, \theta, \tau)\} = \left(\int_0^{+\infty} \int_0^{+\infty} \exp \{ n\alpha \ln(\beta) - k \ln(\Gamma(\alpha)) + \theta\alpha - \tau\beta \} d\beta d\alpha \right)^{-1}.$$

Pelo método de Laplace, temos que:

$$\exp\{A(n, k, \theta, \tau)\} \approx \sqrt{2\pi v_2} \exp \{ L_2(\alpha_2^*, \beta_2^*) \},$$

sendo que, especificamente para este caso, por sorte, há forma analítica aproximada para α_2^*, β_2^* . De fato, veja que $L_2(\alpha, \beta) = n\alpha \ln(\beta) - k \ln(\Gamma(\alpha)) + \theta\alpha - \tau\beta$, daí:

$$\begin{aligned}\frac{\partial}{\partial \alpha} L_2(\alpha, \beta) &= n \ln(\beta) - k\psi(\alpha) + \theta \\ \frac{\partial}{\partial \beta} L_2(\alpha, \beta) &= n \frac{\alpha}{\beta} - \tau,\end{aligned}$$

onde ψ é a função *digamma*.

Da segunda equação obtemos que:

$$\frac{\partial}{\partial \beta} L_2(\alpha^*, \beta^*) = 0 \iff n \frac{\alpha^*}{\beta^*} = \tau \iff \beta^* = \frac{n}{\tau} \alpha^*$$

Substituindo o valor de β^* na primeira equação e usando uma aproximação de primeira ordem para a função *digamma* obtemos que:

$$\begin{aligned}\frac{\partial}{\partial \alpha} L_2(\alpha^*, \beta^*) &= n \ln\left(\frac{n}{\tau}\right) + n \ln(\alpha^*) - k\psi(\alpha^*) + \theta, \\ &\approx n \ln\left(\frac{n}{\tau}\right) + n \ln(\alpha^*) - k \ln(\alpha^*) + \frac{k}{2\alpha^*} + \theta,\end{aligned}$$

Usando o Wolfram, encontramos que $\frac{\partial}{\partial \alpha} L_2(\alpha^*, \beta^*) = 0$ se, e somente se:

$$\alpha^* = \frac{k}{2(k-n)W\left(\frac{k(2^{1-\frac{n}{k}} e^{\theta/k} \frac{n}{\tau}) - k/(k-n)}{k-n}\right)},$$

onde W é a função W de Lambert.

Veja que a solução proposta não está bem definida no caso $n = k$, ademais, como discutido anteriormente, a aproximação de primeira ordem para a função *digamma* deixa muito a desejar. Contudo, no caso onde $n = k$ podemos usar uma aproximação de segunda ordem para a função *digamma*, pois, neste caso, obtemos:

$$\frac{\partial}{\partial \alpha} L_2(\alpha^*, \beta^*) \approx n \ln\left(\frac{n}{\tau}\right) + n \ln(\alpha^*) - k \ln(\alpha^*) + \frac{k}{2\alpha^*} + \frac{k}{12\alpha^{*2}} + \theta = n \ln\left(\frac{n}{\tau}\right) + \frac{n}{2\alpha^*} + \frac{k}{12\alpha^{*2}} + \theta,$$

daí:

$$\frac{\partial}{\partial \alpha} L_2(\alpha^*, \beta^*) = 0 \iff \alpha^* = \frac{1}{3 + \sqrt{9 + 12 \left(\ln\left(\frac{\tau}{n}\right) - \frac{\theta}{n} \right)}}.$$

Observe que, na solução acima, se $\ln\left(\frac{\tau}{n}\right) < \frac{\theta}{n} - \frac{3}{4}$, então α^* não está bem definido, o que é algo problemático. Dito isso, como visto anteriormente, já é necessário que $\ln\left(\frac{\tau}{n}\right) > \frac{\theta}{n}$, a restrição $\ln\left(\frac{\tau}{n}\right) > \frac{\theta}{n} - \frac{3}{4}$ já está sendo satisfeita.

Vale destacar que, se Π é a posteriori depois de se observar uma amostra de tamanho m do modelo observacional ($m \gg 0$), então $\ln\left(\frac{\tau}{n}\right) \approx \ln\left(\frac{1}{m} \sum x_i\right)$ e $\theta/n \approx \frac{1}{m} \sum \ln(x_i)$, daí, como a função \log é concava, vale que $\ln\left(\frac{\tau}{n}\right) > \frac{\theta}{n}$, ou seja, α^* está bem definido, logo, após observar uma quantidade razoável de dados, estaremos livres do risco de que α^* não existir.

Uma vez obtido α^* e β^* , podemos obter uma aproximação para $A(n, k, \theta, \tau)$ e então obter $\mathbb{E}_p[H_p] = \nabla A(n, k, \theta, \tau)$.

Como mencionado anteriormente, também podemos obter $\mathbb{E}_p[H_p]$ por integração numérica e obter um resultado semelhante. A princípio, não seria viável obter $\mathbb{E}_p[H_p]$ por integração numérica, pois a integral que devemos calcular é dupla, inviabilizando os métodos determinísticos que conheço (eles funcionam, mas ficam com o custo computacional irrazoável), ademais, não podemos usar integração por Monte Carlo, pois, para o método de Newton-Raphson, devemos calcular as derivadas de $\mathbb{E}_p[H_p]$ e como não conhecemos a forma analítica de $\mathbb{E}_p[H_p]$, devemos recorrer a diferenciação numérica, porém o erro de Monte Carlo impede que isso seja feito (o ruído aleatório intrínscio ao método de Monte Carlo “ofusca” o valor das derivadas, sendo necessário usar amostras irrazoavelmente grandes para contornar esse problema).

Dito isso, durante a construção deste relatório, observamos que $\mathbb{E}_p[H_p]$ pode ser escrito como uma integral que depende apenas de α , o que viabiliza o uso de métodos numérico determinísticos para calcular $\mathbb{E}_p[H_p]$ (especificamente, usamos Quadratura Gaussiana, que é o método *default* da função *integrate* do *R*), usando essa abordagem obtemos um resultado parecido com o resultado usando a aproximação de Laplace, porém, acredito que o uso de Quadratura Gaussiana seja mais adequado, pois evita o uso de aproximações, gerando assim valores que são numericamente iguais ao verdadeiro.

Vale destacar que, apesar da abordagem usando o método de Laplace não ser a que recomendo, acredito que o desenvolvimento dessa abordagem foi útil para dar alguns *insights* que não obteríamos se tivéssemos feito uso de integração numérica desde o início.

Isso concluí a análise sobre o uso da aproximação de Laplace para calcular $\mathbb{E}_p[H_p]$. Resolvido esta questão, resta apenas usar o algoritmo de Newton-Raphson para resolver o sistema $\mathbb{E}_p[H_p] = \mathbb{E}_q[H_p]$.

Como mencionado anteriormente, há certas escolhas de parâmetros para a distribuição conjugada para as quais Π não é própria, sendo que não conseguimos uma forma de garantir que o algoritmo de Newton-Raphson não passe por parâmetros inadequados. Não bastasse isso, também não conseguimos garantir que o sistema sequer tenha solução.

Tentando aplicar o a metodologia em dados simulados, independente da inicialização do modelo, sempre há alguma iteração onde o algoritmo de Newton-Raphson tem problemas. Ademais também reparamos que há uma distorção considerável dos dados após a compatibilização das priors, de modo que tivemos problemas mesmo nas ocasiões em que o algoritmo de Newton-Raphson de fato convergiu.

Por último, apresentarei na próxima sessão uma abordagem alternativa que permite simplificar a resolução do sistema e obter forma analítica aproximada para $\mathbb{E}_p[H_p]$.

Proposta alternativa: Aproximação da distribuição marginal de α

Lembremos que, se $\alpha, \beta \sim \Pi(n, k, \tau, \theta)$, então $\beta|\alpha \sim \mathcal{G}(n\alpha + 1, \tau)$ e:

$$f(\alpha) \propto \frac{\Gamma(n\alpha + 1)}{\Gamma(\alpha)^k} \exp\{(\theta - n \ln(\tau))\alpha\}.$$

Ademais, a fórmula de Stirling nos diz que:

$$\Gamma(x + 1) \approx \sqrt{2\pi x} \left(\frac{x}{e}\right)^x.$$

Usando essa aproximação na densidade de α obtemos que:

$$\begin{aligned}
f(\alpha) &\propto \frac{\Gamma(n\alpha + 1)}{\Gamma(\alpha)^k} \exp\{(\theta - n \ln(\tau))\alpha\} \\
&= \frac{\Gamma(n\alpha + 1)}{\alpha^{-k} \Gamma(\alpha + 1)^k} \exp\{(\theta - n \ln(\tau))\alpha\} \\
&\approx \frac{\sqrt{2\pi n} \alpha^{\frac{1}{2}} n^{n\alpha} \alpha^{n\alpha} e^{-n\alpha}}{\sqrt{2\pi}^k \alpha^{-k} \alpha^{\frac{k}{2}} \alpha^{k\alpha} e^{-k\alpha}} \exp\{(\theta - n \ln(\tau))\alpha\} \\
&= \frac{\sqrt{n}}{(2\pi)^{\frac{k-1}{2}}} \alpha^{\frac{1}{2} + \frac{k}{2}} \alpha^{(n-k)\alpha} e^{-(n-k)\alpha} \exp\{(\theta - n \ln(\tau/n))\alpha\}.
\end{aligned}$$

Até aqui a equação acima não é particularmente útil, porém, se assumirmos $n = k$, obtemos:

$$f(\alpha) \propto \alpha^{\frac{1}{2} + \frac{n}{2}} \exp\{(\theta - n \ln(\tau/n))\alpha\}.$$

Ao olhar com carinho, podemos reparar que α tem distribuição aproximada $\mathcal{G}\left(\frac{n+3}{2}, n \ln(\tau/n) - \theta\right)$ (e esta é uma boa aproximação). Com essa informação, temos que:

$$\mathbb{E}_p[\alpha] \approx \frac{n+3}{2(n \ln(\tau/n) - \theta)}$$

O que pode ser usado para resolver mais uma das equações do sistema $\mathbb{E}_p[H_p] = \mathbb{E}_q[H_p]$ sem recorrer a métodos numéricos.

Para obter solução analítica para todo o sistema resta apenas calcular um valor esperado, uma vez que, ao supor $n = k$ reduzimos o sistema para 3 equações, das quais temos solução analítica para duas. Se conseguirmos resolver a última de forma analítica, podemos conseguir uma forma garantida de fazer a compatibilização das prioris sem depender de métodos iterativos.

O valor esperado que resta calcular é $\mathbb{E}_p[\alpha \ln(\beta) - \ln(\Gamma(\alpha))] = \mathbb{E}_p[\alpha \psi(n\alpha + 1) - \ln(\tau)\alpha - \ln(\Gamma(\alpha))]$ (note que, ao supor $n = k$, H_p se tornou outro vetor). Usando que:

$$\begin{aligned}
\psi(n\alpha + 1) &\approx \psi(n\alpha) \approx \ln(n\alpha) - \frac{1}{2n\alpha} - \frac{1}{12n^2\alpha^2} \\
\ln(\Gamma(\alpha)) &\approx \alpha \ln(\alpha) - \frac{1}{2} \ln(\alpha) + \frac{1}{12\alpha},
\end{aligned}$$

podemos obter a seguinte aproximação:

$$\begin{aligned}
\alpha \psi(n\alpha + 1) - \ln(\tau)\alpha - \ln(\Gamma(\alpha)) &\approx \alpha \ln(n\alpha) - \frac{1}{2n} - \frac{1}{12n^2\alpha} - \alpha \ln(\alpha) + \frac{1}{2} \ln(\alpha) - \frac{1}{12\alpha} \\
&= \alpha \ln(n) - \left(\frac{1}{2n} - \frac{1}{12n^2} - \frac{1}{12}\right) \frac{1}{\alpha} + \frac{1}{2} \ln(\alpha).
\end{aligned}$$

Por último, considerando que $\mathbb{E}_p[\frac{1}{\alpha}] \approx \frac{2(n \ln(\tau/n) - \theta)}{n+1}$ e $\mathbb{E}_p[\ln(\alpha)] \approx \psi\left(\frac{n+3}{2}\right) - \ln(n \ln(\tau/n) - \theta)$ (pois α tem distribuição aproximada $\mathcal{G}\left(\frac{n+3}{2}, n \ln(\tau/n) - \theta\right)$), então:

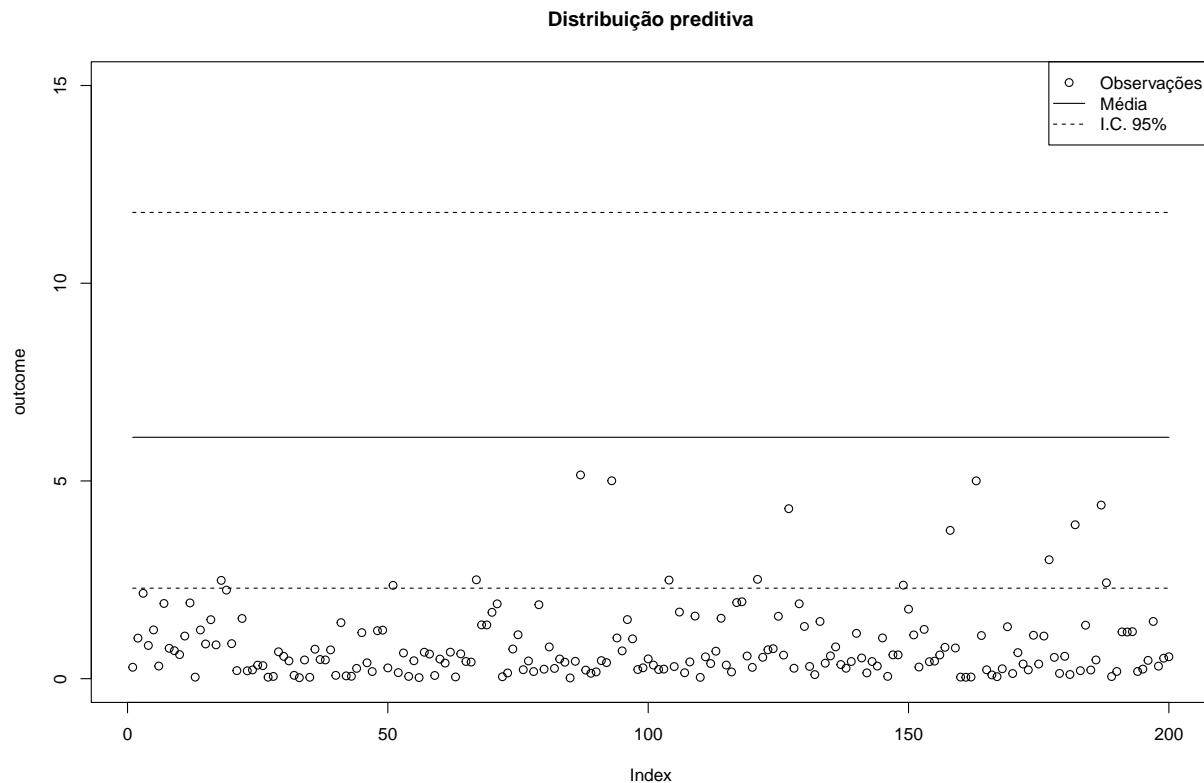
$$\mathbb{E}_p[\alpha \ln(\beta) - \ln(\Gamma(\alpha))] \approx \frac{n+3}{2(n \ln(\tau/n) - \theta)} \ln(n) - \left(\frac{1}{2n} - \frac{1}{12n^2} - \frac{1}{12}\right) \frac{2(n \ln(\tau/n) - \theta)}{n+1} + \frac{1}{2} \psi\left(\frac{n+3}{2}\right) - \ln(n \ln(\tau/n) - \theta).$$

Usando a equação acima, ainda temos de usar Newton-Raphson para resolver o sistema, porém com estas simplificações o algoritmo ficou muito mais rápido e converge sem dificuldades, possibilitando a

compatibilização das prioris. Vale destacar também que a aproximação usada é muito boa, de modo que não há perdas significativas na resolução do sistema (pelo menos nos casos em que testei).

Usando os resultados acima, foi possível fazer o ajuste do GDLM, prossigamos então a análise apresentando o resultado do ajuste do modelo usando as aproximações propostas.

Para testar a qualidade do ajuste, geramos uma amostra i.i.d. com 200 elementos da distribuição $\mathcal{G}(1, 1)$ e usamos a abordagem do artigo k-paramétrico para estimar os parâmetros α e β da distribuição dos dados. O resultado pode ser observado a seguir:



Parâmetro alpha

Valor real: 1

Média: 0.468725817021833

Mediana: 0.404045061851186

Quantil de 0.975: 1.19424744613269

Quantil de 0.025: 0.136368395122457

Parâmetro beta

Valor real: 1

Média: 0.0718832090036268

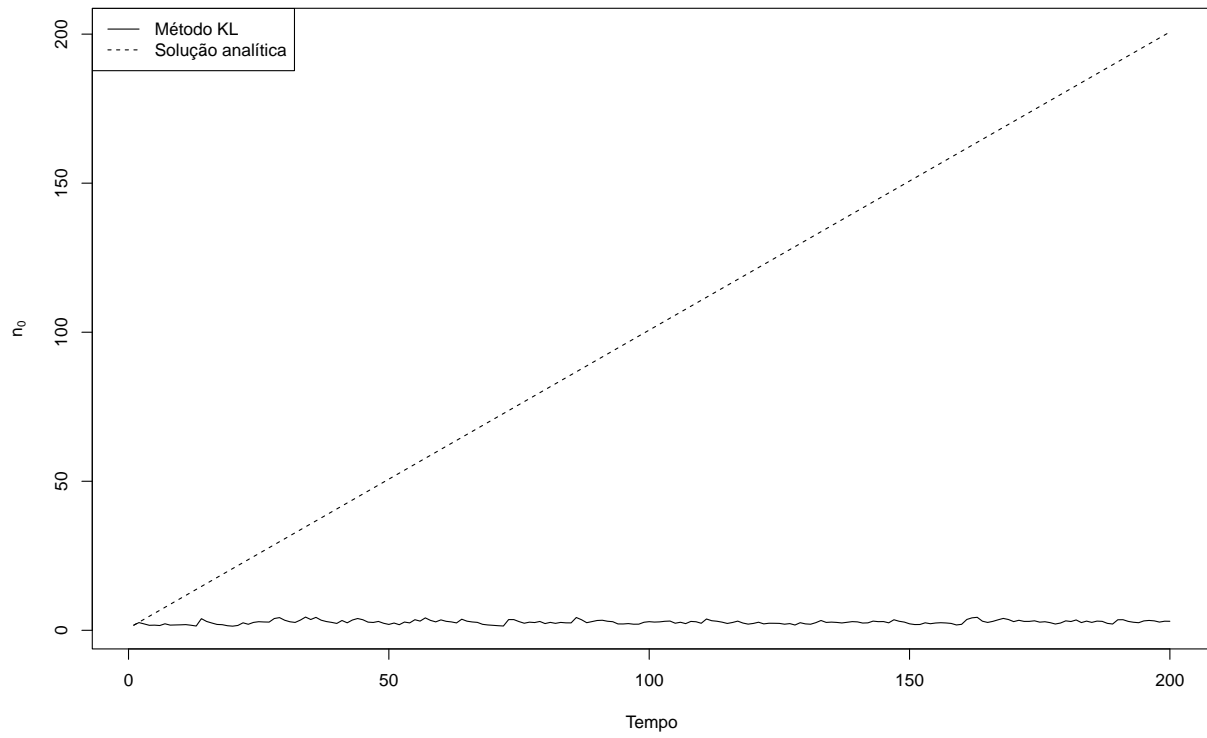
Mediana: 0.0673312685816263

Quantil de 0.975: 0.137111834199079

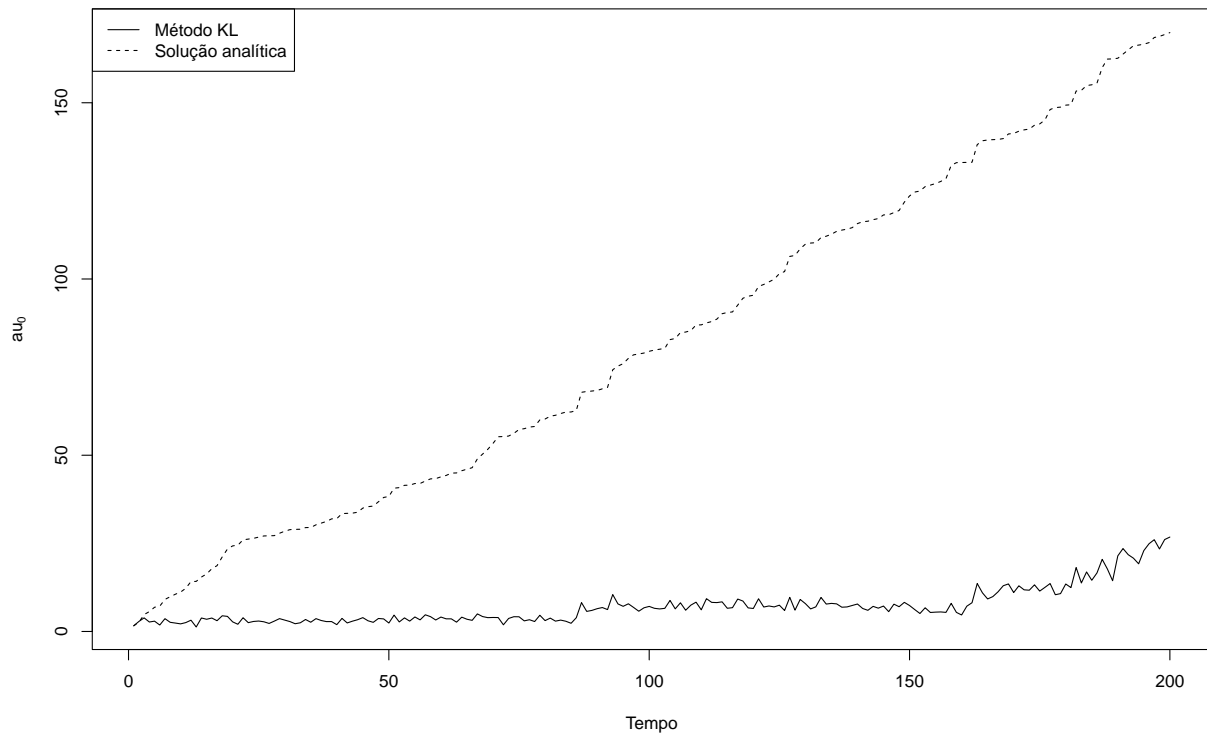
Quantil de 0.025: 0.0329446389131496

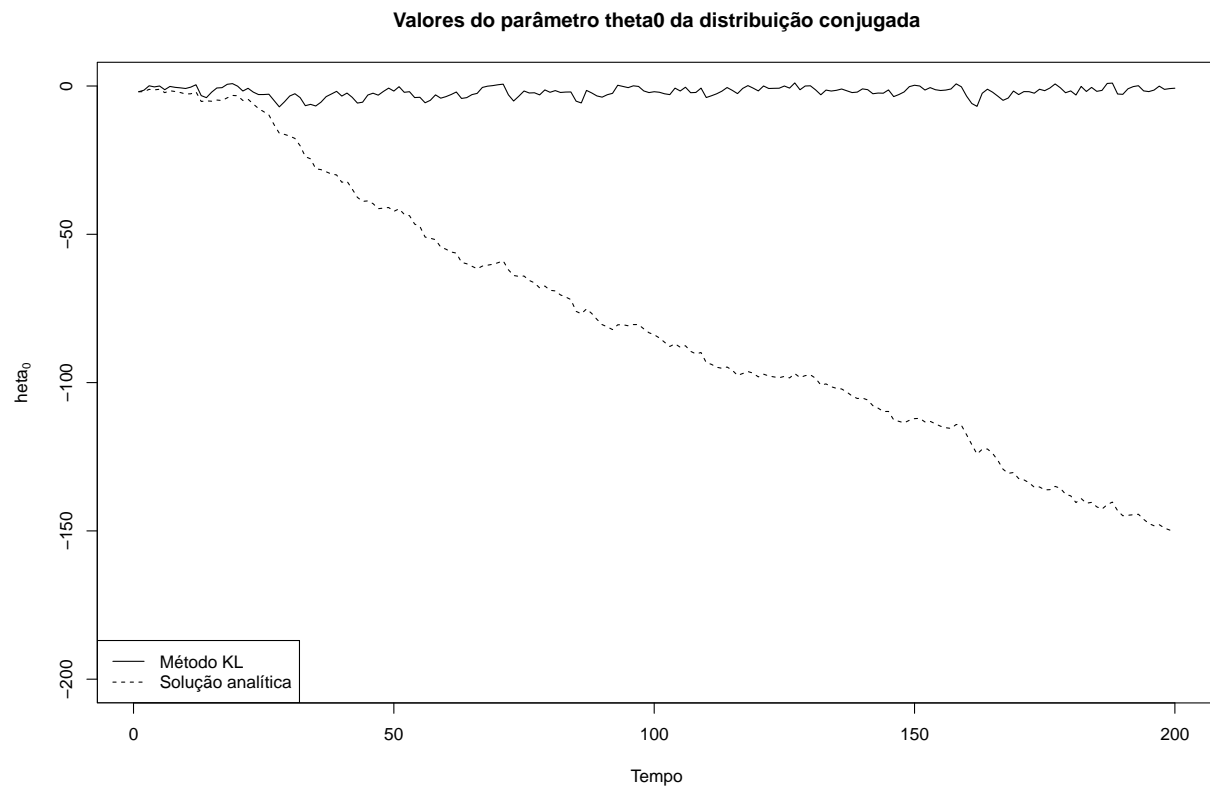
Podemos observar que a estimativa dos parâmetros está bem ruim. Afim de tentar entender a causa para essas estimativas ruins, apresentamos um gráfico com a estimação dos parâmetros a cada iteração:

Valores do parâmetro n_0 da distribuição conjugada



Valores do parâmetro τ_0 da distribuição conjugada





Claramente temos um problema na compatibilização das prioris, especificamente, parece que a informação adquirida após se observar o dado é perdida durante a compatibilização. Mais investigações serão feitas para tentar resolver esse problema.