

Relatório sobre a qualidade das aproximações

Silvaneio Viera dos Santos Junior

2022-10-05

Introdução

Neste relato apresento os resultados dos experimentos que fiz para testar a qualidade das aproximações usadas no artigo sobre GDLM k-paramétricos.

Na primeira seção, será apresentado o problema sobre a qualidade da aproximação de primeira ordem da função digamma (ψ) e duas aproximações alternativas com precisão satisfatória. No final da análise concluímos que é necessário o uso de ordem superior a 1, desta forma, usaremos a aproximação de segunda ordem para a função *digamma* sempre que necessário.

Na seção seguinte, discutiremos a qualidade da aproximação usada para o modelo com resposta Poisson para diversos valores do parâmetro de taxa (ou média). Apresentaremos também exemplos de ajustes para valores “extremos” do parâmetro de taxa afim de encontrar um conjunto onde o uso do modelo é sempre válido. Com as análises desta seção, concluímos que o ajuste é bom no caso Poisson com média superior a 1 e razoável para Poisson’s com média acima de 0.75.

Na 3ª seção abordaremos a qualidade da aproximação para o modelo com resposta Gamma com parâmetro de forma (ϕ) conhecido. De forma análoga ao caso Poisson, avaliaremos o ajuste para diversos valores do parâmetro de forma (a média do dado não afeta a qualidade do ajuste) e encontramos um conjunto onde o ajuste é sempre válido. No final desta análise, concluímos que o ajuste do modelo é bom para os casos Gamma com ϕ superior a 0.75 e satisfatório para casos Gamma com ϕ superior a 0.5.

Na 4ª seção discutiremos a qualidade da aproximação usada para o modelo com resposta Multinomial para diversos valores dos parâmetros N e p . Para este caso, avaliaremos especificamente o caso binomial, tendo em mente que o resultado da análise se estende aos casos com mais de 2 categorias. Ao final da análise, concluímos que o ajuste é bom para uma certa categoria da multinomial se o valor esperado de observações para esta categoria for maior ou igual a 0.2.

Por último, discutiremos a qualidade da aproximação para o modelo com resposta Normal com variância desconhecida. Este é o caso mais delicado, pois a qualidade do ajuste **não** depende da distribuição dos dados (mostraremos o porquê), desta forma, este é o caso onde o ajuste do modelo se mostra mais discrepante com a solução analítica. Posto isto, o ajuste ainda me parece satisfatório, ademais, é possível estabelecer condições para que a compatibilização das distribuições seja válida (estas condições dependem da estrutura do modelo, mas não da distribuição do dado). Alternativamente, também apresento uma proposta que melhora a qualidade do ajuste e não impõe nenhuma restrições na estrutura do modelo.

Qualidade da aproximação para a função digamma

Ao montar os sistemas que devem ser resolvidos para obter a priori conjugada, frequentemente nos deparamos com a função digamma (ψ). Por exemplo, no caso do modelo com reposta Poisson, adotamos uma priori Gamma; para encontrar os parâmetros da distribuição Gamma que melhor aproxima a distribuição do preditor linear (que é assumida como normal), devemos encontrar $\tau_1 = \alpha$ e $\tau_2 = \beta$ tais que:

$$\begin{aligned}\psi(\alpha) - \ln(\beta) &= f_t \\ \frac{\alpha}{\beta} &= \exp\{f_t + q_t/2\},\end{aligned}\tag{1}$$

onde f_t é a média do preditor linear e q_t é a sua variância.

Para resolver esse sistema sem recorrer a método numéricos, usamos que:

$$\psi(x) = \ln(x) - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \dots$$

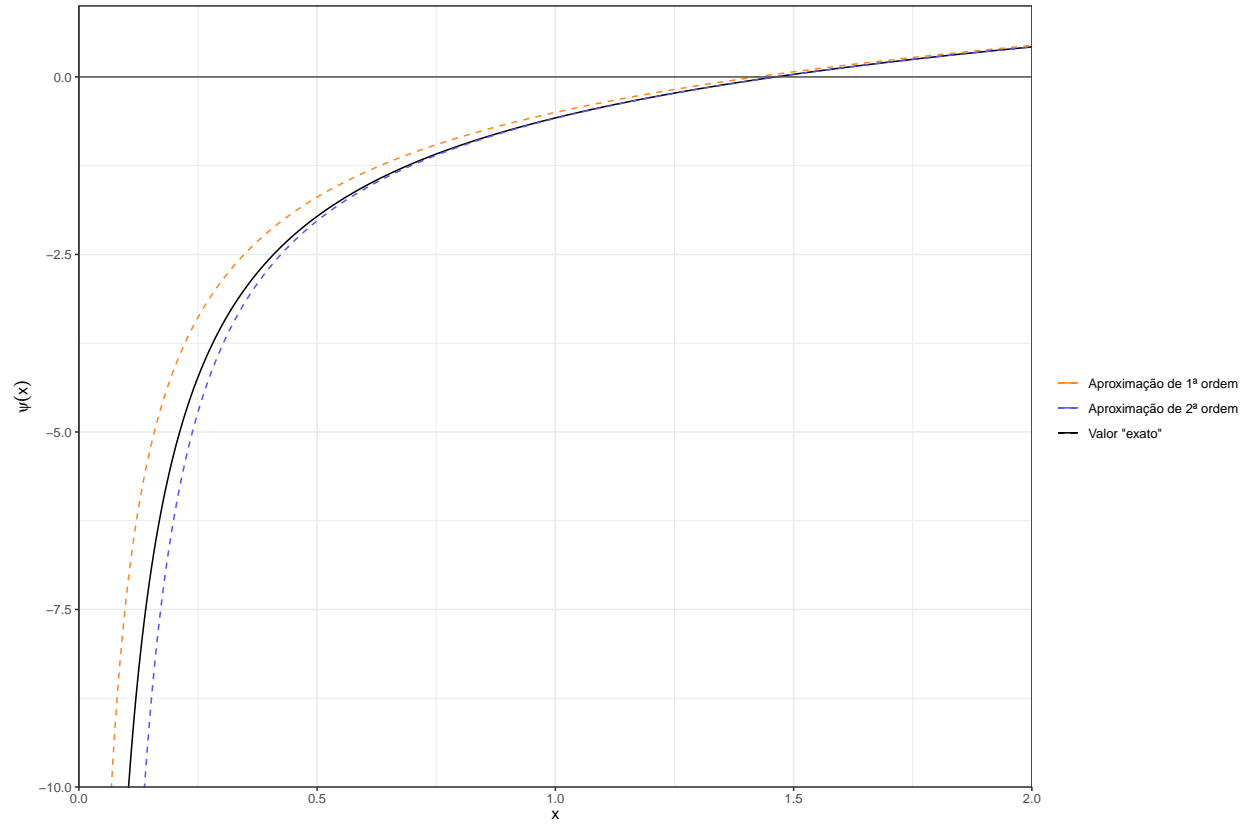
Assim, uma aproximação de primeira ordem para $\psi(x)$ seria $\tilde{\psi}_1(x) = \ln(x) - \frac{1}{2x}$. Com esta aproximação, o sistema (1) tem a seguinte solução:

$$\begin{aligned}\alpha &= q_t^{-1} \\ \beta &= \alpha \exp\{-f_t - q_t/2\}\end{aligned}$$

De forma análoga, uma aproximação de segunda ordem para $\psi(x)$ seria $\tilde{\psi}_2(x) = \ln(x) - \frac{1}{2x} - \frac{1}{12x^2}$. Com esta aproximação, o sistema (1) tem a seguinte solução:

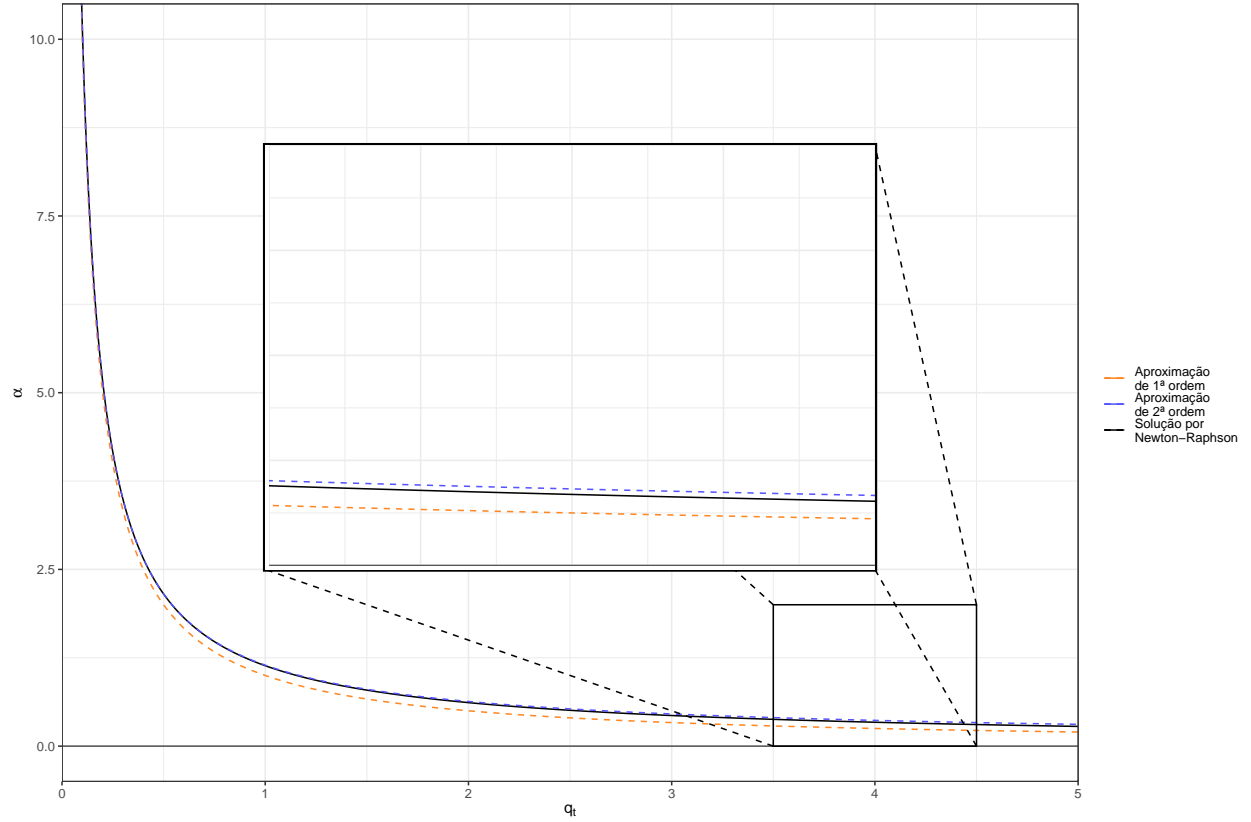
$$\begin{aligned}\alpha &= \frac{1}{-3 + 3\sqrt{1 + \frac{2}{3}q_t}} \\ \beta &= \alpha \exp\{-f_t - q_t/2\}\end{aligned}$$

No gráfico a seguir, apresentamos uma comparação entre as aproximações de primeira e segunda ordem e tomando como valor “exato” de ψ os valores fornecidos pela função *digamma* do R:



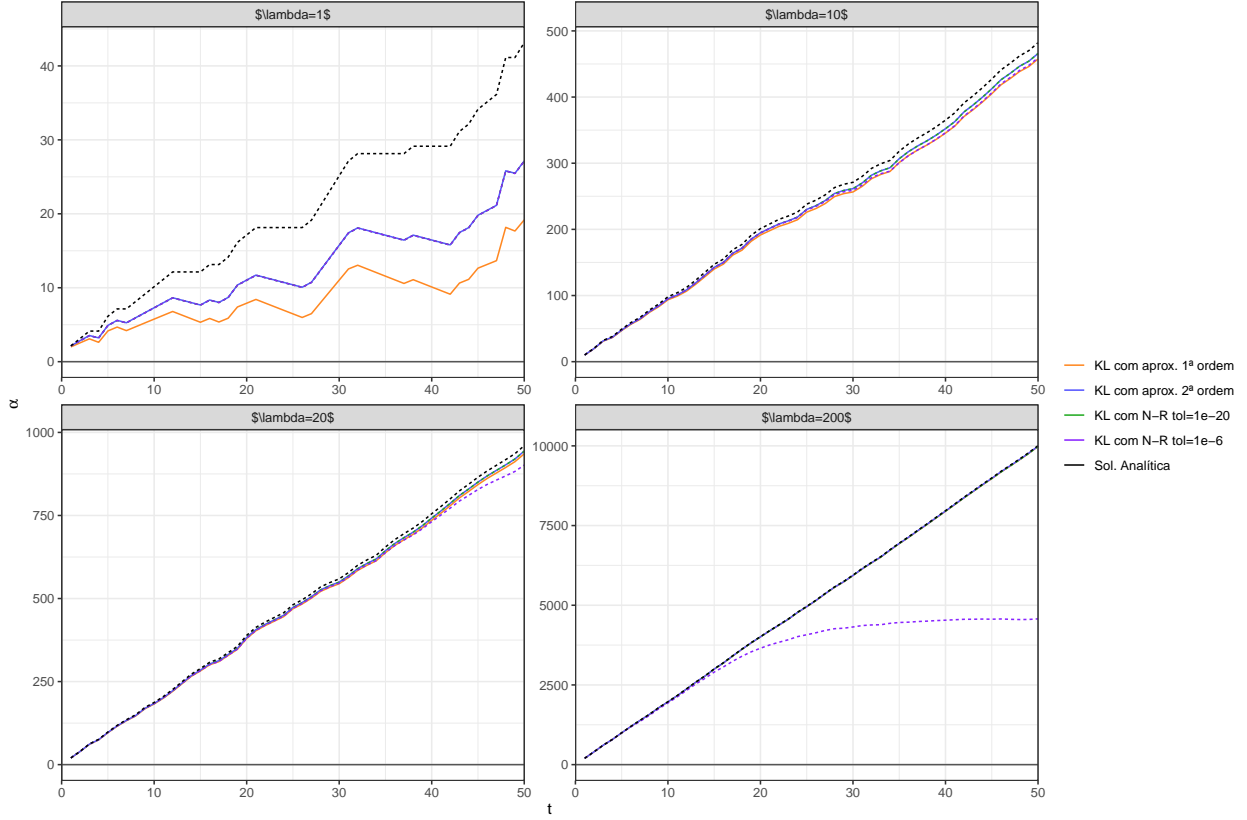
Veja que, para valores de x inferiores a 1, a aproximação de primeira ordem deixa muito a desejar. Já a aproximação de segunda ordem, apesar de também se afastar do valor “exato”, tem uma precisão muito maior.

A seguir apresentamos uma comparação entre as soluções obtidas por cada aproximação para o sistema (1) usando como valor de referência a solução obtida com o método de Newton-Raphson para o sistema original com tolerância de $1e - 20$ (isto é, a solução obtida está na vizinhança de $1e - 20$ da solução exata):



Observe que, para valores suficientemente pequenos de q_t , as duas soluções são praticamente iguais, mas para valores maiores de q_t a aproximação de 1ª ordem deixa muito a desejar, enquanto a aproximação de 2ª ordem é bem próxima da solução de referência.

Adiante apresentaremos o resultado do ajuste de um modelo dinâmico Poisson para cada uma das aproximações. Os ajustes foram feitos com amostras i.i.d. de Poisson's com taxas 1, 10, 20 e 200. Para cada amostra, ajustamos um modelo dinâmico Poisson com estrutura de primeira ordem para o nível e com fator de desconto igual a 1. Neste cenário, temos solução analítica para a estimação dos parâmetros, assim, podemos comparar os parâmetros estimados pelo modelo com a solução analítica. Como referência para a qualidade da aproximação da função ψ , também ajustamos 2 modelos usando o método de Newton-Raphson para resolver o sistema (1). Inicialmente, calculamos as soluções com Newton-Raphson usando tolerância de $1e - 6$ para a solução, porém este valor se mostrou inadequado, o que nos levou a fazer um ajuste alternativo com tolerância de $1e - 20$ para a solução. Por último, no gráfico a seguir, vamos nos referir a abordagem proposta no artigo sobre GDLM k-paramétricos como KL (ou aproximação KL, em alguns casos), sendo que vamos nos referir aos casos onde essa abordagem foi usada junto com o método de Newton-Raphson para a resolução do sistema (1) como KL com N-R.



Observe que a solução KL N-R com tolerância $1e - 6$ deixa bastante a desejar quando α é grande, sendo necessário maior precisão numérica para a solução. A solução KL N-R com tolerância $1e - 20$, apesar de boa, tem um custo computacional elevado, sendo necessárias várias iterações do algoritmo para obter uma solução com a precisão desejada. Em contra-partida, a solução KL com aproximação de 2ª produz um resultado essencialmente igual ao KL N-R com tolerância $1e - 20$, mas sem a necessidade de métodos numéricos para a solução. Vale observar que, apesar das soluções encontradas para λ pequeno serem consideravelmente diferentes da solução analítica, veremos na próxima seção que esta discrepância no espaço paramétrico **não** se traduz em uma grande discrepância de ajuste do modelo.

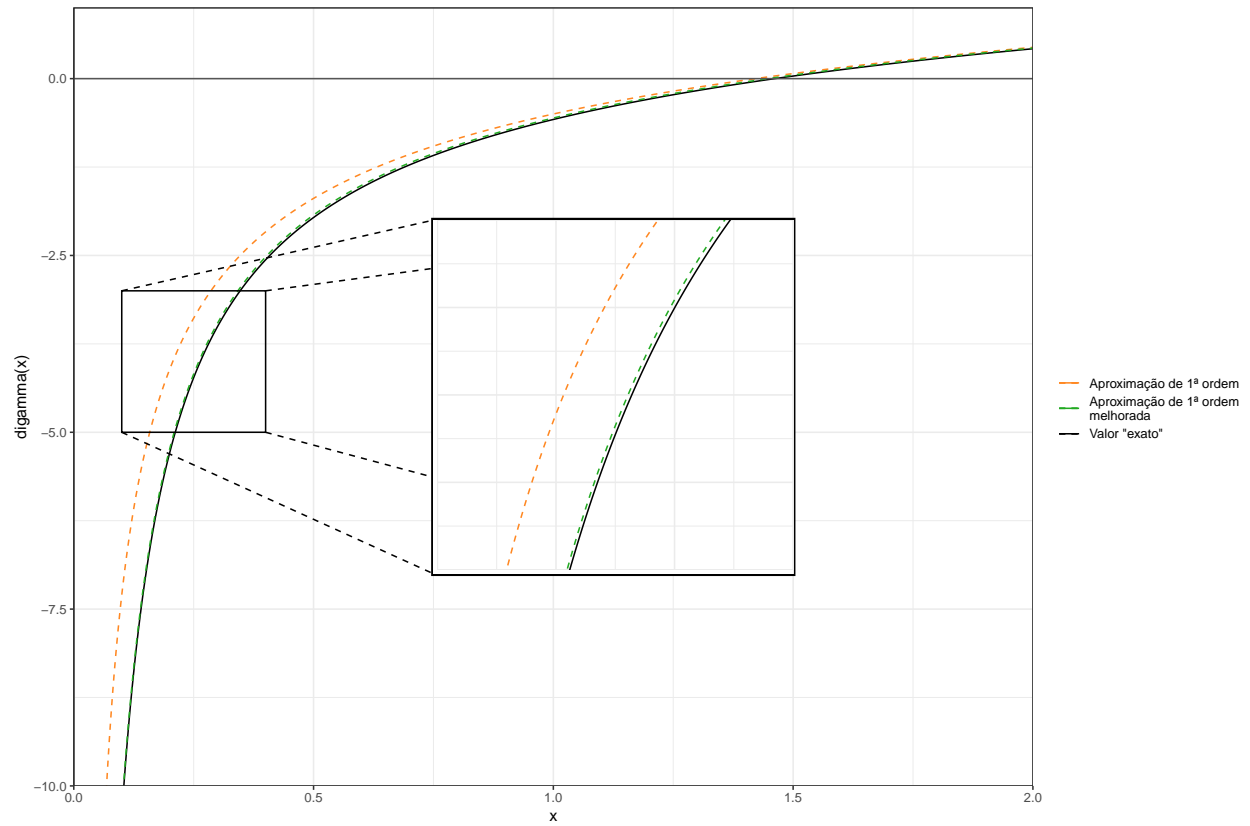
Por último, antes de finalizar esta seção, acredito que seja válido discutir uma possível alternativa para a aproximação de 2ª ordem de ψ . Sabemos que a função ψ tem a seguinte propriedade:

$$\psi(x) = \psi(x+1) - \frac{1}{x}$$

Como aproximação de 1ª ordem é boa para valores de x maiores do que 1, podemos usar uma aproximação de primeira ordem para $\psi(x+1)$ e encontrar uma aproximação muito boa para ψ , de fato, a aproximação obtida é:

$$\psi(x) \approx \ln(x+1) - \frac{1}{2(x+1)} - \frac{1}{x}$$

No gráfico a seguir comparamos a aproximação de 1ª ordem original com a aproximação melhorada:



Veja que esta aproximação é muito melhor do que a aproximação de 2ª ordem! Porém, esta aproximação ajuda em muito pouco na resolução do sistema (1), ademais, como já vimos, não há ganho significativo em melhorar a aproximação de segunda ordem uma vez que a solução obtida com esta aproximação é comparável à solução obtida por N-R com tolerância $1e - 20$.

Qualidade da compatibilização de priori: Caso Poisson

Primeiramente, vamos apresentar modelo que estamos usando para fixar a notação. Sejam y uma v.a. tal que:

$$\begin{aligned} y|\eta &\sim \text{Poisson}(\eta) \\ \ln(\eta) &= \lambda \\ \lambda &\sim \mathcal{N}(f, q). \end{aligned}$$

Nesta notação, η é o parâmetro de taxa da Poisson e λ é o preditor linear.

Na abordagem KL, para obter a priori conjugada para η ($\eta \sim \mathcal{G}(\tau_1, \tau_2)$), devemos resolver o seguinte sistema:

$$\begin{aligned} \psi(\tau_1) - \ln(\tau_2) &= f \\ \frac{\tau_1}{\tau_2} &= \exp\{f + q/2\}, \end{aligned} \tag{2}$$

Usando a aproximação de segunda ordem para a função ψ , obtemos a seguinte solução para o sistema:

$$\begin{aligned} \tau_1 &= \frac{1}{-3 + 3\sqrt{1 + \frac{2}{3}q}} \\ \tau_2 &= \tau_1 \exp\{-f - q/2\}, \end{aligned} \tag{3}$$

Obtidos τ_1 e τ_2 , definimos $\tau_1^* = \tau_1 + y$ e $\tau_2^* = \tau_2 + 1$ de modo que $\eta|y \sim \mathcal{G}(\tau_1^*, \tau_2^*)$

Para obter a posteriori Normal para λ ($\lambda|y \sim \mathcal{N}(f^*, q^*)$), devemos tomar f^* e q^* tais que:

$$\begin{aligned} f^* &= \psi(\tau_1^*) - \ln(\tau_2^*) \\ q^* &= \psi'(\tau_1^*). \end{aligned} \tag{4}$$

Uma dúvida a ser levantada é se a aproximação das distribuições usando a metodologia proposta não induz algum tipo de “distorção” na estimação. Uma forma de se averiguar isso seria observar o que ocorre com os parâmetros da Normal após a compatibilização quando não atualizamos a distribuição de η , isto é, se começamos com uma distribuição $\mathcal{N}(f, q)$ para λ , obtemos a distribuição conjugada $\mathcal{G}(\tau_1, \tau_2)$ para η e depois obtemos uma nova distribuição $\mathcal{N}(f^*, q^*)$ para λ sem atualizar a distribuição de η , o quanto $\mathcal{N}(f, q)$ difere de $\mathcal{N}(f^*, q^*)$?

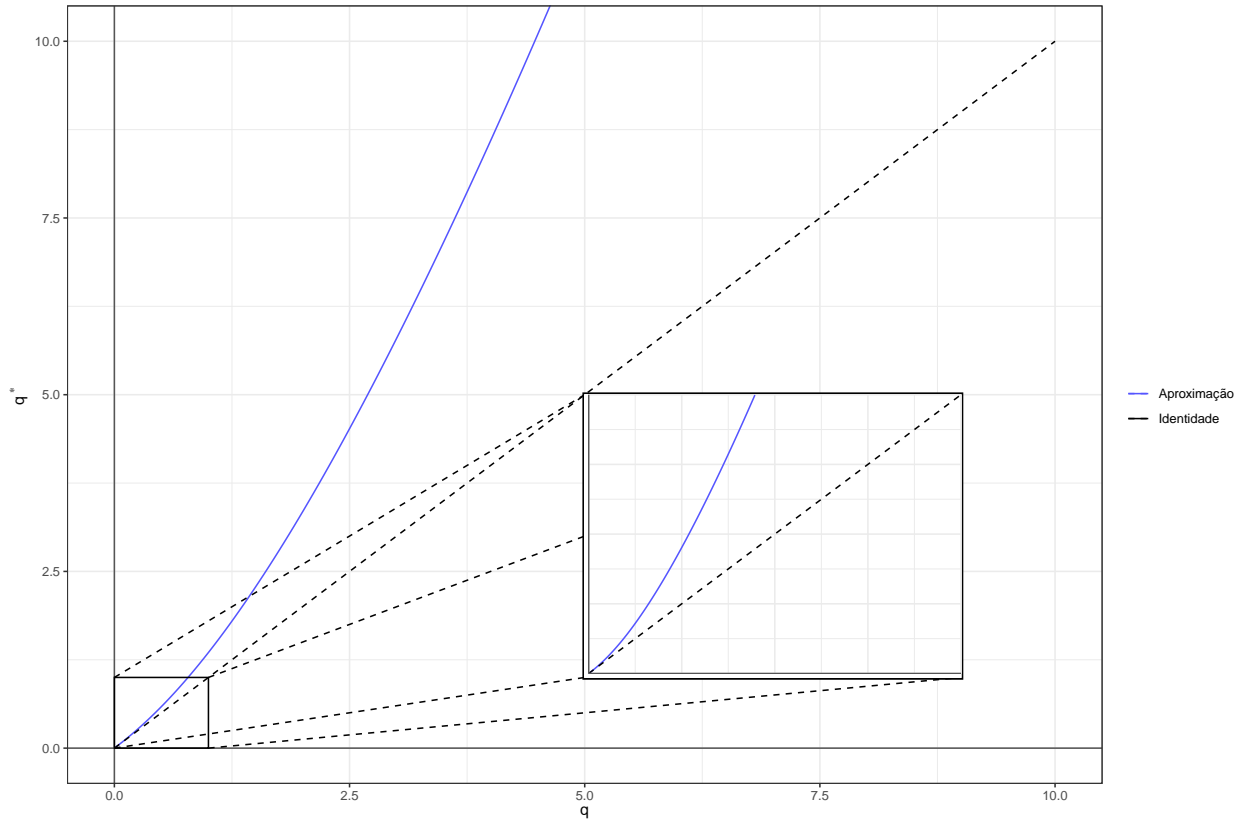
Para responder esta pergunta podemos observar que:

$$f^* = \psi(\tau_1^*) - \ln(\tau_2^*) = \psi(\tau_1) - \ln(\tau_2) = f.$$

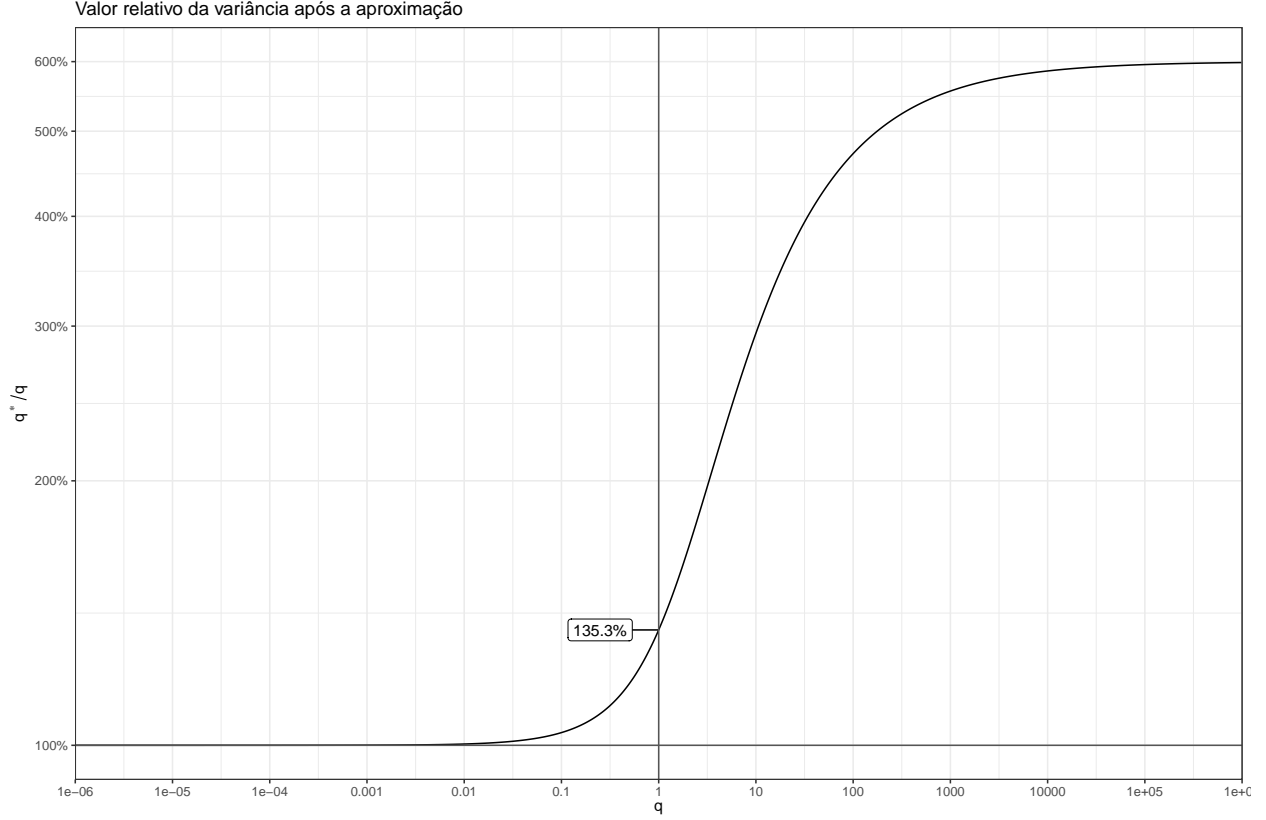
Logo, a menos de desvios devidos a aproximação de ψ , as duas normais tem a mesma média. Para o valor de q^* , observe que:

$$q^* = \psi'(\tau_1^*) = \psi'(\tau_1) = \psi' \left(\frac{1}{-3 + 3\sqrt{1 + \frac{2}{3}q}} \right).$$

Avaliaremos a seguir o gráfico de q contra q^* para verificar se há alguma distorção do valor original de q após passar pelas aproximações:



Observe que $q^* > q$, ademais a discrepância entre q e q^* **parece** crescer exponencialmente. Para tentar entender o comportamento desta aproximação, especialmente nos casos onde $q \approx 0$ e $q \approx +\infty$, podemos olhar para a razão $\frac{q^*}{q}$ contra q na escala log x log:



Veja que o aumento percentual da variância converge para 500% quando $q \rightarrow +\infty$. Se por um lado 500% é um valor bem alto, pelo menos esta razão não diverge para $+\infty$. Vale destacar também que este aumento na variância é razoavelmente alto mesmo para valores usuais de q , por exemplo, para $q = 1$, temos um aumento de 35.3% na variância da distribuição aproximada.

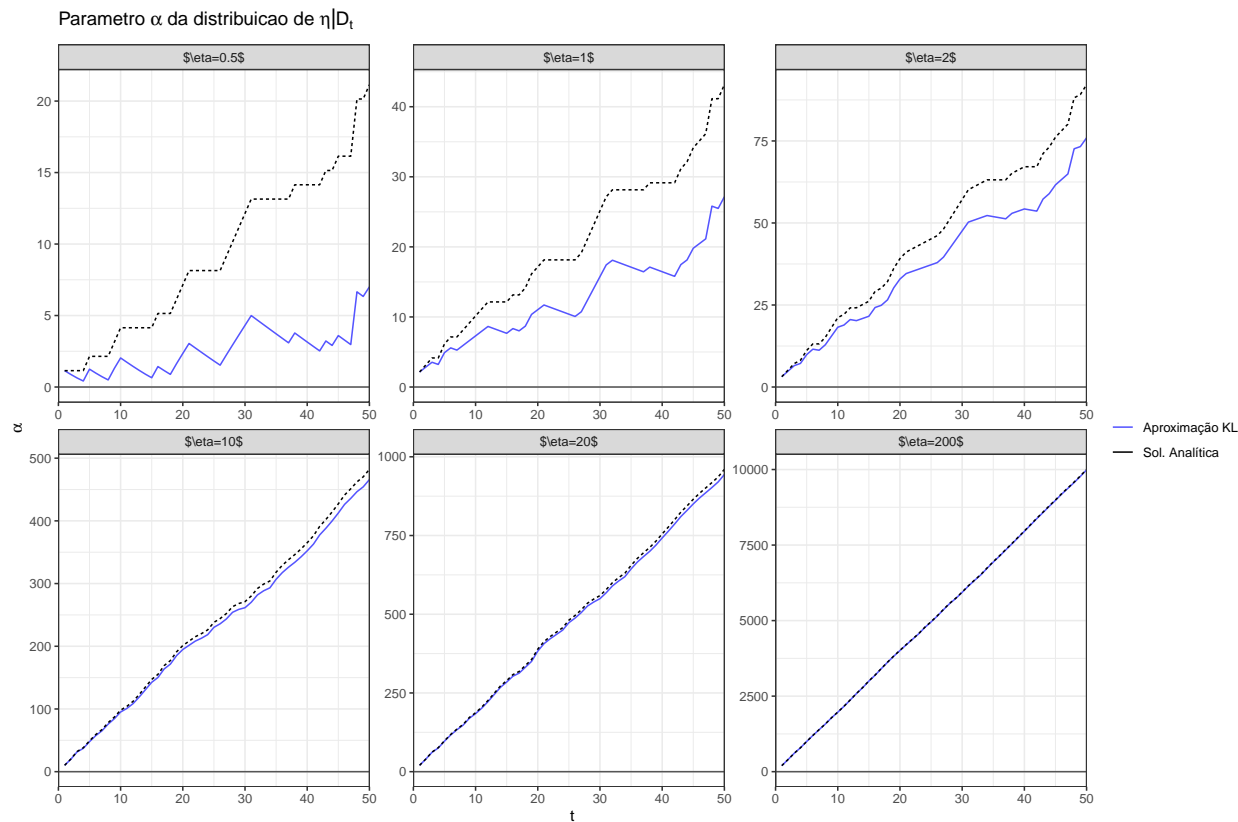
Vale notar que q é inversamente proporcional à τ_1 , desta forma, a aproximação ser boa para valores pequenos q é o mesmo que dizer que a aproximação é boa para valores grande de τ_1 . Desta forma a qualidade da aproximação depende de que τ_1 seja suficientemente grande.

Na prática, a dependência no valor de τ_1 é pouco impeditiva, pois a cada iteração do algoritmo devemos atualizar τ_1 para $\tau_1^* = \tau_1 + y$, assim, mesmo que comecemos com um valor de τ_1 baixo, após algumas iterações teremos uma aproximação suficientemente boa para que a abordagem funcione bem. Dito isso, ainda há perda de informação na transição entre as distribuições. Suponha que estamos no caso onde, a cada iteração, perdemos exatamente 80% o valor de τ_1 no processo de compatibilização, ao obter $\tau_1^* = \tau_1 + y$ aproveitaremos na próxima iteração apenas 20% do valor observado de y , isso vai fazer com que o valor de τ_1 cresça numa velocidade inferior ao que seria “exato”. Contudo, se y for suficientemente grande, mesmo aproveitando apenas 20% do valor observado, a atualização pode “deslocar” o parâmetro τ_1 para uma região onde a perda de informação seja pequena. Assim, se a média de y é suficientemente grande, o ajuste ficará bom para qualquer τ_1 inicial. Resta então a seguinte questão: Qual o menor valor médio de y que garanta um bom ajuste do modelo? O restante desta seção será dedicado a tentar responder esta pergunta.

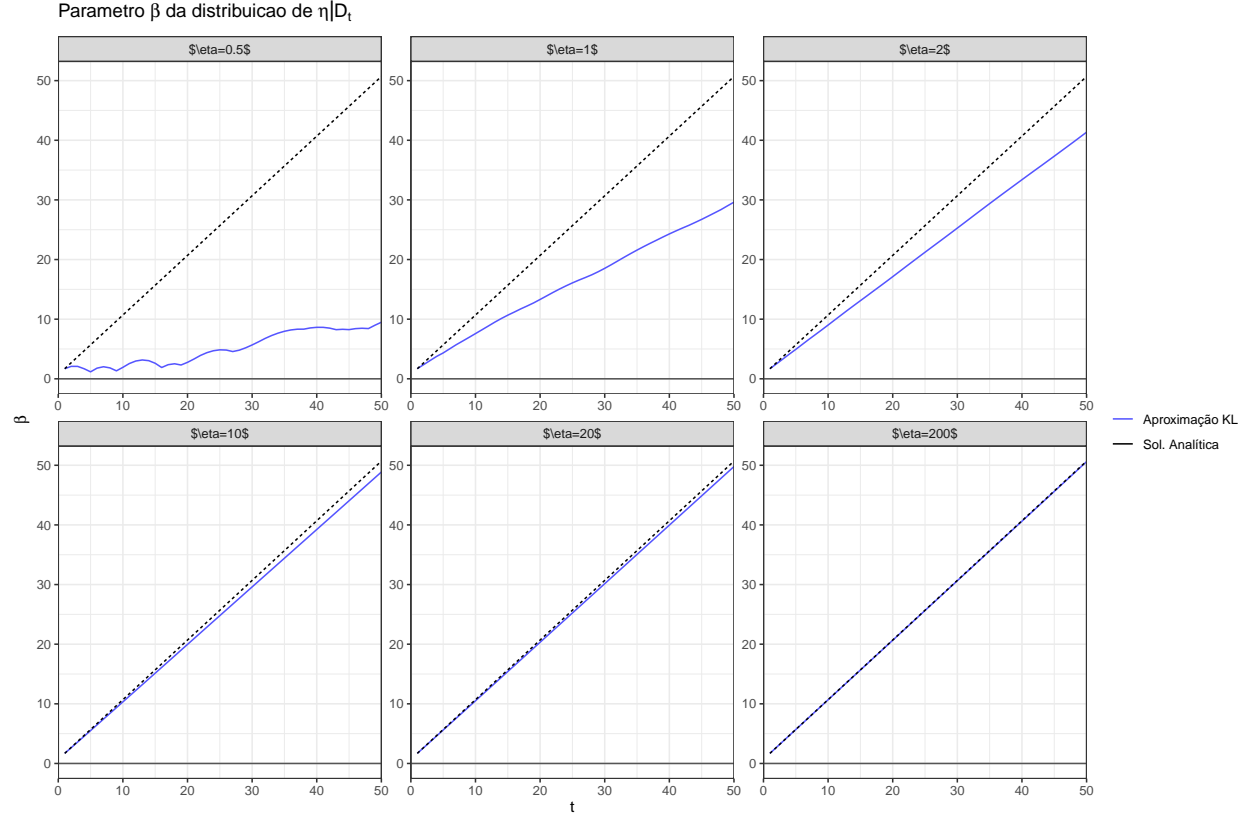
Primeiro teste: Utilizando amostras i.i.d. de distribuições Poisson com diversas médias, vamos comparar a estimativa dos parâmetros fornecida pela abordagem KL com o valor analítico dos parâmetros. Para cada tempo t , temos que a distribuição a posteriori analítica para η é:

$$\begin{aligned} \eta | \mathcal{D}_t &\sim \mathcal{G}(\alpha_t, \beta_t) \\ \alpha_t &= \alpha_{t-1} + y_t \\ \beta_t &= \beta_{t-1} + 1, \end{aligned}$$

sendo que α_0 e β_0 foram escolhidos de modo que a solução analítica e a abordagem KL tenham a mesma priori.



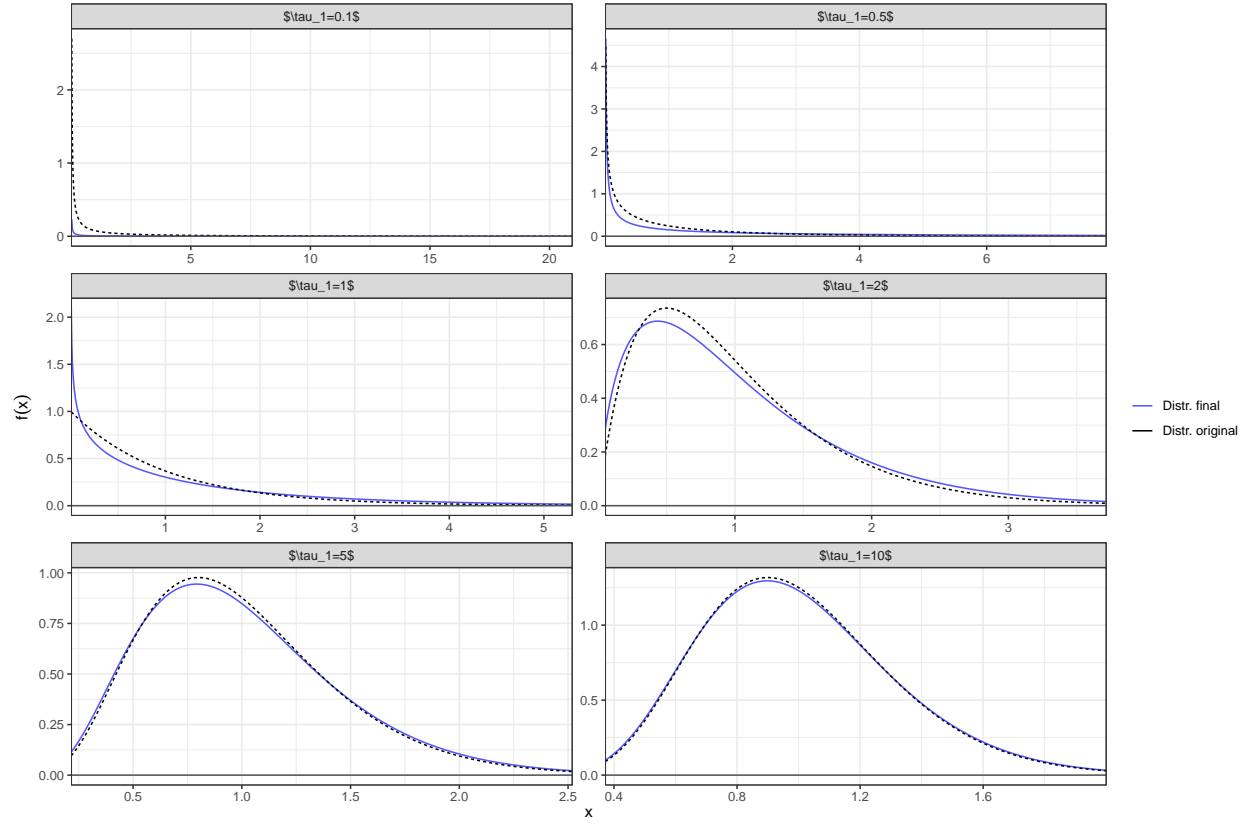
Observe no gráfico acima que, quando a linha tracejada (sol. analítica) se mantém estável (isto é $y_i = 0$), o valor de α diminui. Isso faz com que a presença de 0's na amostra prejudique a qualidade do ajuste. Dito isso, se a frequência com que 0's ocorrem é baixa, então não há problemas.



No gráfico acima podemos ver um resultado análogo a estimação de α .

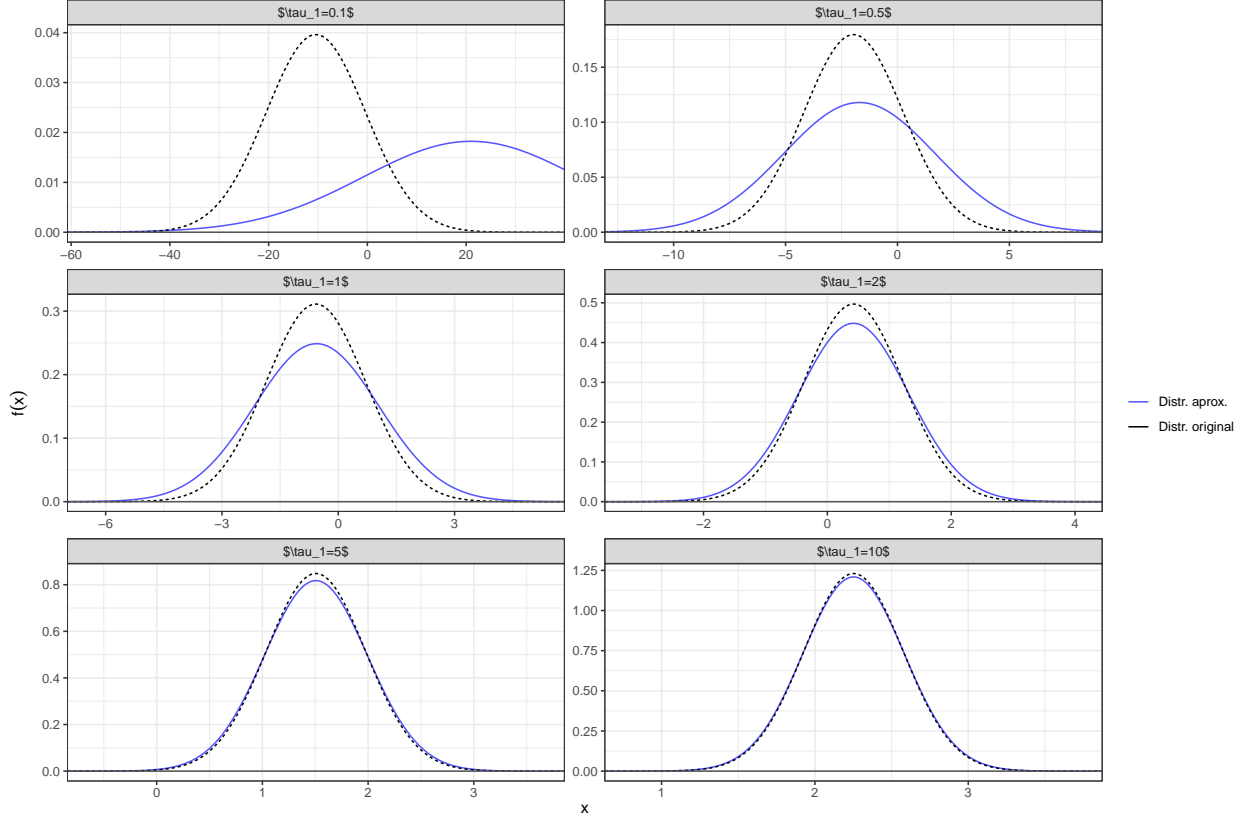
Observe que, a partir de $\eta = 10$ temos que a aproximação KL gera resultado bem próximos da solução analítica. Dito isso, vale destacar que estamos olhando a distância entre a aproximação KL e a solução analítica no espaço paramétrico, mas uma grande distância no espaço paramétrico não necessariamente implica que as distribuições são de fato muito diferentes.

A seguir vamos exibir comparações entre a distribuição de η antes e depois da compatibilização para diversos valores de $\alpha = \tau_1$. Em cada um dos casos apresentado a frente, começamos com $\eta \sim \mathcal{G}(\tau_1, \tau_2)$, depois obtemos a distribuição aproximada Normal para o λ e por fim usamos a distribuição de λ para encontrar uma nova distribuição $\mathcal{G}(\tau_1^*, \tau_2^*)$ para η . Avaliamos então a diferença entre a distribuição original e a distribuição final de η :



Veja que a aproximação já parece bem razoável para $\tau_1 = 2$ e me parece “decente” para $\tau_1 = 1$. Já para valores de $\tau_1 < 1$, a aproximação não parece tão confiável.

Adiante exibimos uma comparação parecida, porém vamos comparar a distribuição de λ . Neste caso, começamos com a distribuição $\lambda \sim \mathcal{N}(f, q)$, obtemos a distribuição aproximada para $\eta \sim \mathcal{G}(\tau_1, \tau_2)$ e, usando a distribuição de η , obtemos uma nova distribuição aproximada para $\lambda \sim \mathcal{N}(f^*, q^*)$. Feito este processo, comparamos a distribuição original de λ com a distribuição final:



Observe que começamos a ter problemas com a média da distribuição quando τ_1 é muito pequeno (como por exemplo $\tau_1 = 0.1$), mas, novamente, a aproximação parece razoável para $\tau_1 \geq 1$.

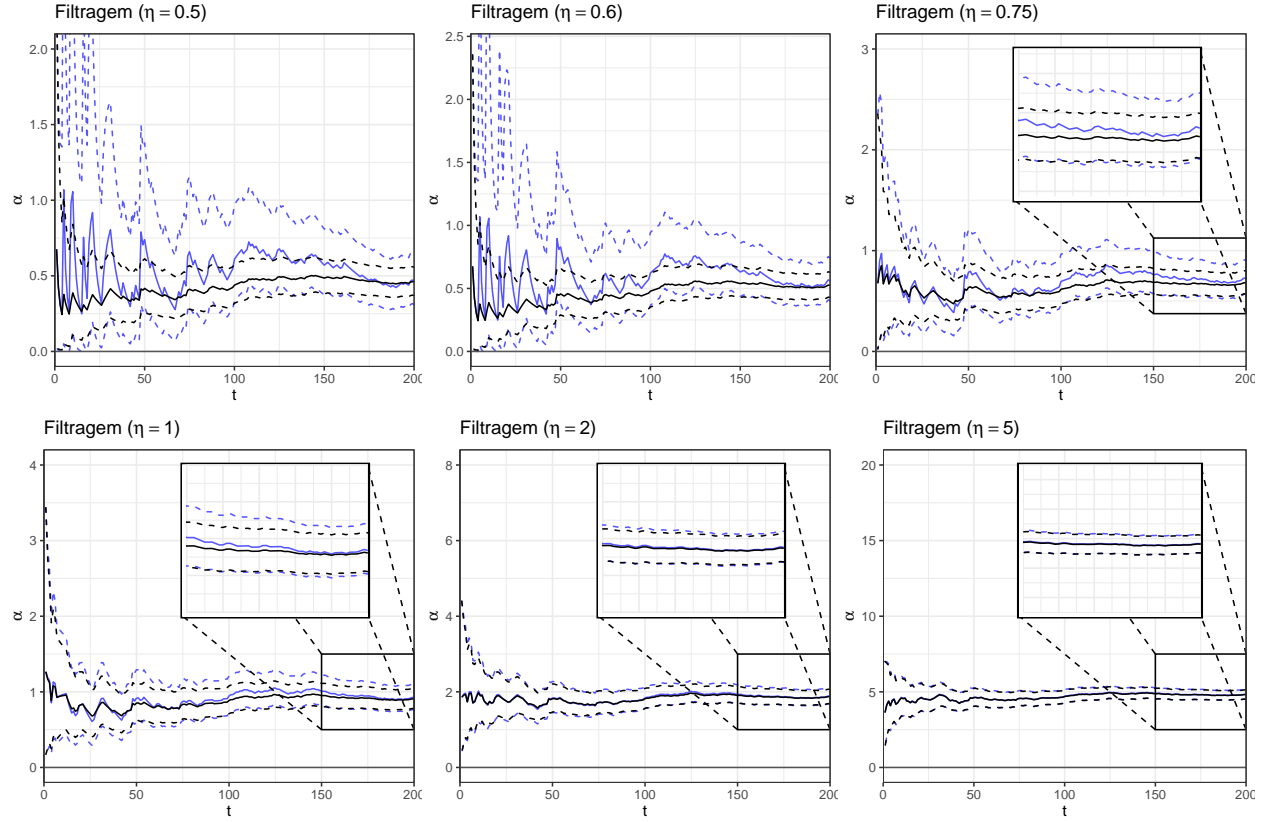
A partir dos gráficos acima, temos indícios de que o ajuste do modelo é válido para Poisson's com taxa pelo menos 1, sendo irrazoável para Poisson's com taxas muito baixas.

Dito isso, a análise se baseia em apenas uma iteração da compatibilização, sendo assim, para avaliar o comportamento de diversas iterações consecutivas da compatibilização, faremos o ajuste com dados simulados.

Nos gráficos a seguir, mostraremos o ajuste feito de forma analítica e com a aproximação KL para amostras i.i.d. de Poisson's com diversas médias. Assim como no primeiro teste que fizemos, a solução analítica para o processo de filtragem de η é dada por:

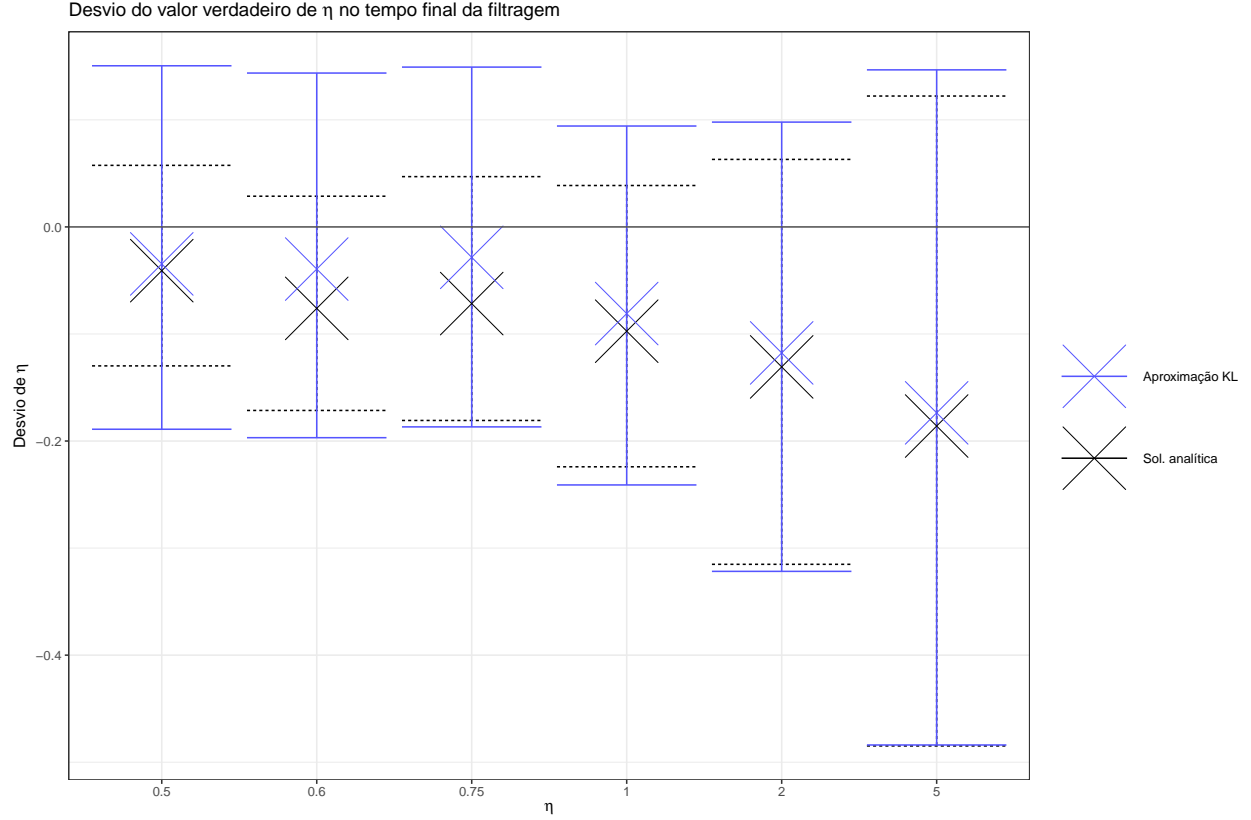
$$\begin{aligned}\eta|\mathcal{D}_t &\sim \mathcal{G}(\alpha_t, \beta_t) \\ \alpha_t &= \alpha_{t-1} + y_t \\ \beta_t &= \beta_{t-1} + 1,\end{aligned}$$

sendo que α_0 e β_0 foram escolhidos de modo que a solução analítica e a abordagem KL tenham a mesma priori.



As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Para a filtragem, temos um resultado satisfatório a partir de $\phi \geq 0.75$ e um resultado muito bom a partir de $\phi \geq 1$.



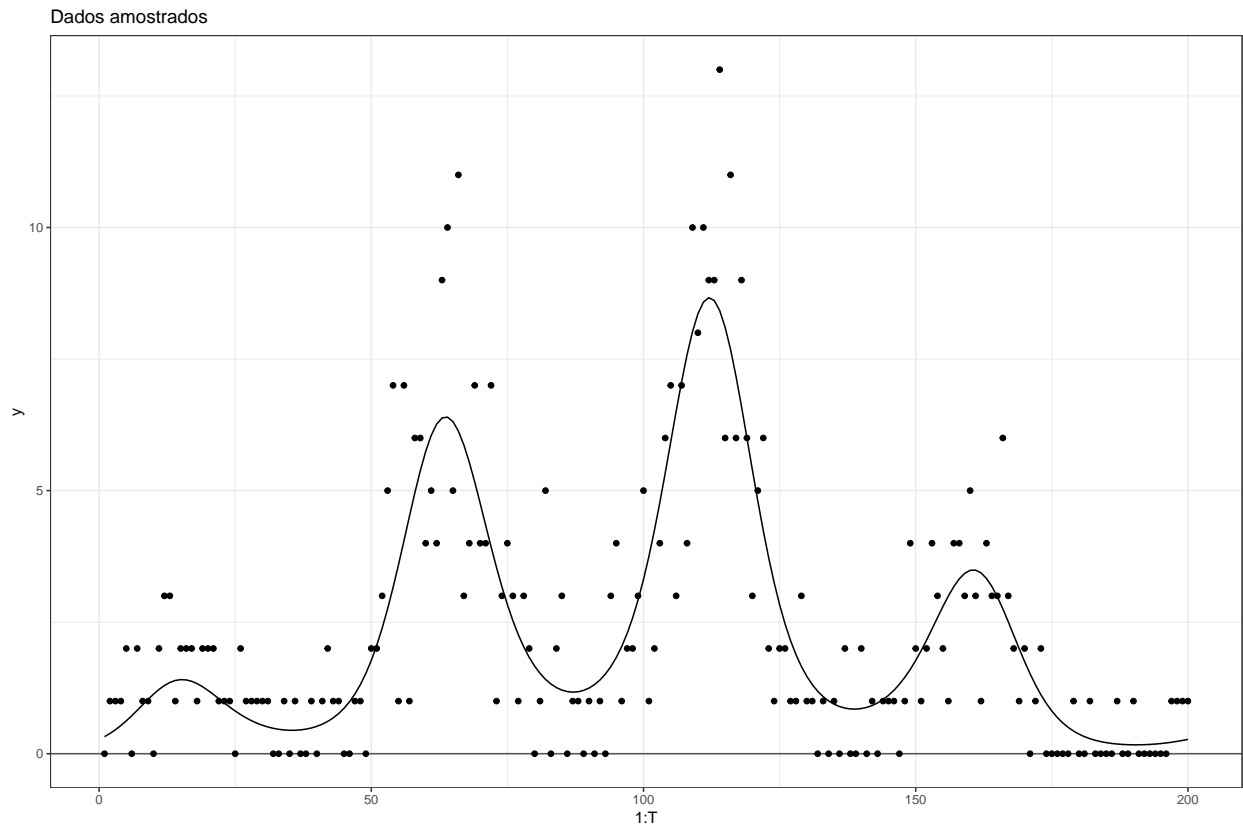
A partir dos gráficos acima, podemos observar que a aproximação KL é satisfatoriamente próxima da solução analítica a partir de $\eta = 1$, sendo que, em geral, a aproximação KL é mais “conservadora”, no sentido de que a amplitude dos intervalos de credibilidade é maior do que a solução analítica (não necessariamente o i.c. da solução analítica está contido no i.c. da aproximação KL). Vale destacar que, como τ_1 está sendo subestimado, temos que a aproximação KL é mais sensível ao dado mais recente, dando um peso maior do que o analítico para o dado recém observado, dito isso, para η razoavelmente grande, essa diferença na estimação da média é mínima.

Ainda no contexto de qualidade de ajuste, os casos em que analisamos até o momento não são realistas, no sentido de que, se há forma analítica de se estimar os parâmetros do modelo, não há necessidade de usar aproximações. Tendo isso em mente, adiante exibiremos o ajuste em um caso onde não há solução analítica (o valor de η tem dinâmica temporal e assume valores inferior a 1 em alguns momentos).

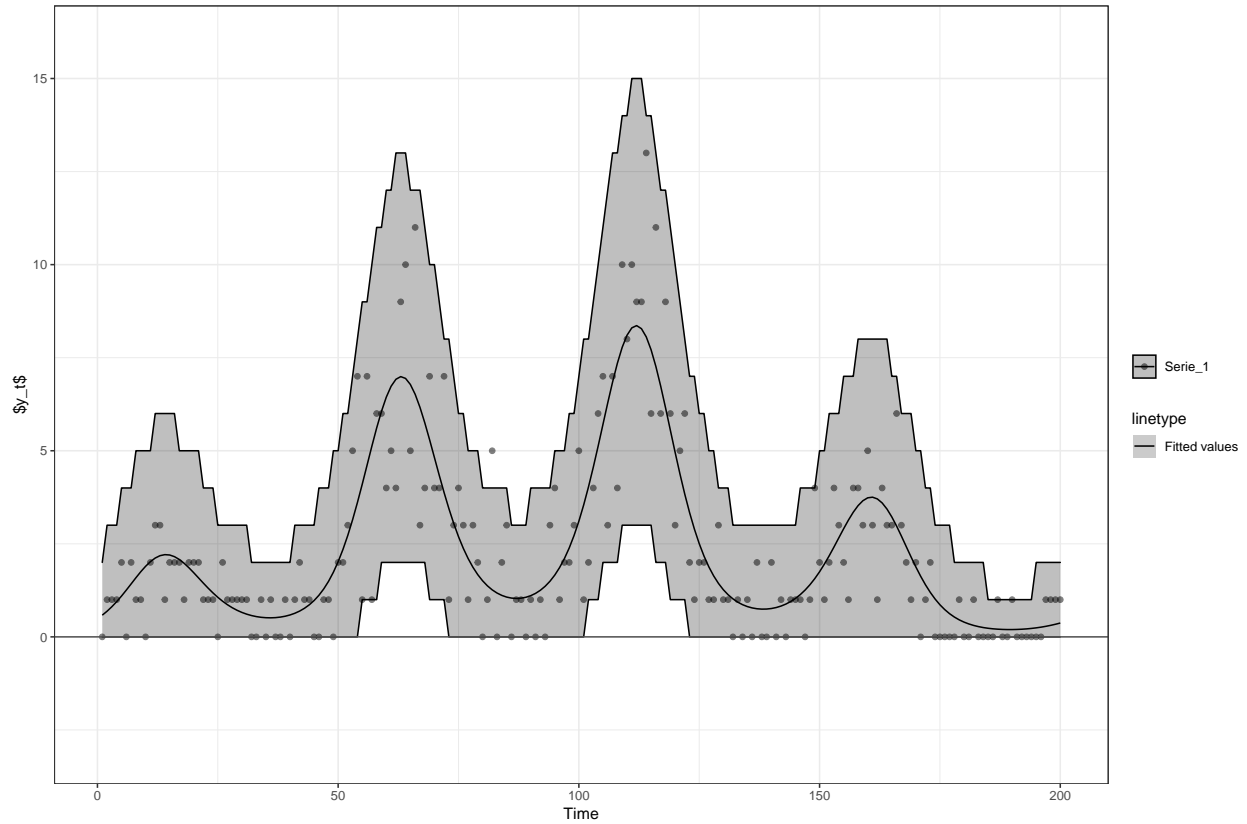
Amostraremos y_1, \dots, y_{200} com as distribuições a seguir:

$$y_i \sim \text{Poisson} \left(\frac{1}{10} \exp \left\{ \sin \left(\frac{i}{25} \pi \right) - \frac{i}{20} \left(\frac{i}{200} - 1 \right) + 1 \right\} \right)$$

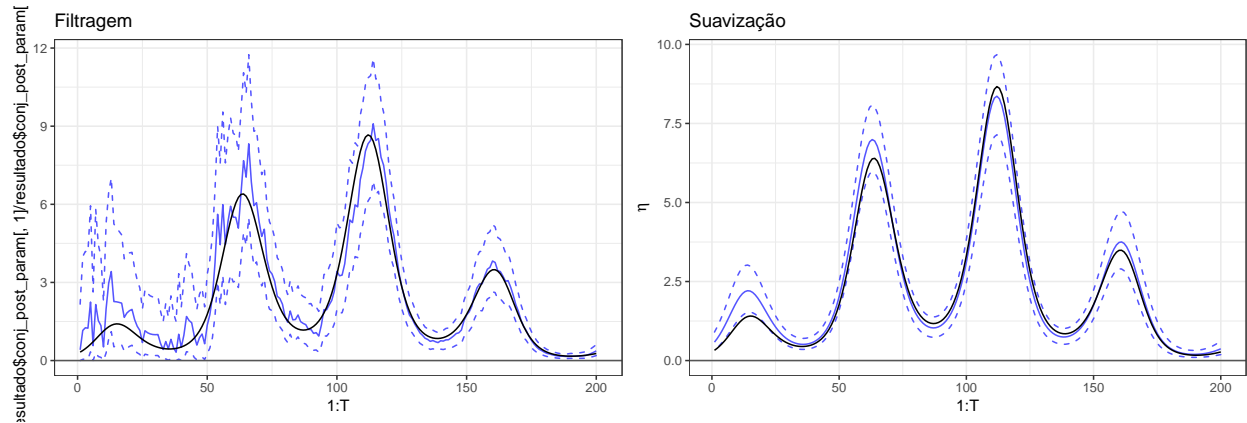
Adiante exibimos o gráfico de η ao longo do tempo junto com os valores amostrados:



Adiante, exibimos o ajuste do modelo (com estrutura compatível com os dados, i.e., um bloco harmonico para a sazonalidade e estrutura de 3ª ordem para o nível):



No gráfico a seguir, comparamos o valor verdadeiro com o valor estimado de η (durante a filtragem e após a suavização):



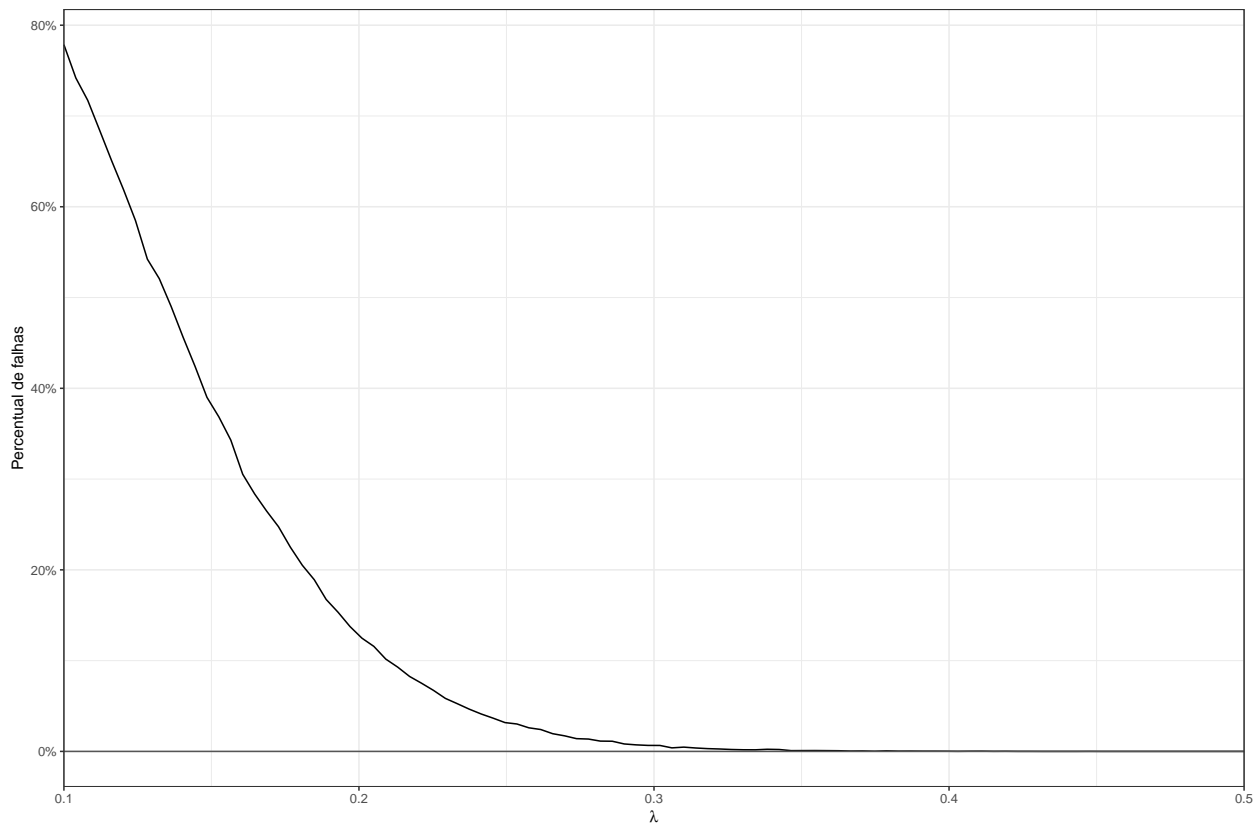
As curvas em preto representam o valor real do parâmetro η a cada tempo e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Podemos observar que o modelo captura bem o comportamento dos dados e estima de forma satisfatória os valores de η ao longo do tempo.

De modo geral, acredito que podemos concluir que a aproximação KL funciona muito bem para o ajuste de dados Poisson cuja média seja igual ou superior a 1 na maior parte do tempo (o que acredito que engloba a maior parte das aplicações). Olhando na perspectiva do pacote para o ajuste do modelo, posso implementar *warnings* ou erros que avisam o usuário quando a η está muito baixo e a estimação dos parâmetros foi comprometida.

Por último, para finalizar esta seção, fiz alguns testes para verificar o qual baixo η pode ser antes de que o ajuste apresente erros numéricos. Estes erros que me refiro são **paralelos** a qualidade da estimação e ocorrem quando há uma sequência muito longa de 0's nas observações, o que leva o preditor linear a divergir para $-\infty$. Vale destacar que a quantidade de 0's consecutivos em uma amostra é aleatória, desta forma, fixado uma Poisson com taxa qualquer, há uma probabilidade positiva de que uma amostra gerada por essa Poisson cause erro numérico no ajuste. Naturalmente, para η grande (> 0.5 , por exemplo), essa probabilidade é praticamente nula, mas foi verificado que para η pequeno essa probabilidade é inaceitavelmente alta (para $\lambda = 0.1$, por exemplo, a probabilidade de falha é de quase 80%).

Para verificar qual o limite para valores aceitáveis de η , rodei diversos modelos com várias taxas (de 0.1 a 0.5) e registrei o percentual de modelos onde ocorreu erro numérico. O gráfico a seguir mostra o resultado do ajuste de 2.000.000 modelos para 100 valores de η (20.000 ajustes para cada η , cada ajuste com uma amostra de 200 elementos).



Veja que para $\eta > 0.4$ o percentual de falha já é suficientemente baixo (menos de 1 em 10.000) e para $\eta > 0.43$ não foram observadas nenhuma falha. Como a qualidade da aproximação KL não parece razoável para valores tão baixo de η , temos que esse problema de longas sequências de 0's na amostra não é relevante, uma vez que só acontece em casos onde o ajuste do modelo não será permitido.

Qualidade da compatibilização de priori: Caso Gamma

O caso Gamma é bem semelhante ao caso Poisson, uma vez que a priori conjugada para a média do modelo Gamma com parâmetro de forma (ϕ) conhecido é Inversa-Gamma. Para fixar a notação, apresentamos a seguir o modelo usado no caso Gamma. Sejam y uma v.a. tal que:

$$\begin{aligned} y|\phi, \mu &\sim \text{Gamma}\left(\phi, \frac{\phi}{\mu}\right) \\ \ln(\mu) &= \lambda \\ \lambda &\sim \mathcal{N}(f, q). \end{aligned}$$

Nesta notação, μ é a média da Gamma e λ é o preditor linear.

Na abordagem KL, para obter a priori conjugada para μ ($\mu \sim \mathcal{IG}(\tau_1, \tau_2)$), devemos resolver o seguinte sistema:

$$\begin{aligned} \psi(\tau_1) - \ln(\tau_2) &= f \\ \frac{\tau_1}{\tau_2} &= \exp\{-f + q/2\}, \end{aligned} \tag{5}$$

Usando a aproximação de segunda ordem para a função ψ , obtemos a seguinte solução para o sistema:

$$\begin{aligned} \tau_1 &= \frac{1}{-3 + 3\sqrt{1 + \frac{2}{3}q}} \\ \tau_2 &= \tau_1 \exp\{f - q/2\} \end{aligned} \tag{6}$$

Obtidos τ_1 e τ_2 , vamos definir $\tau_1^* = \tau_1 + \phi$ e $\tau_2^* = \tau_2 + \phi y$ de modo que $\mu|y \sim \mathcal{G}(\tau_1^*, \tau_2^*)$

Para obter a posteriori Normal para λ ($\lambda|y \sim \mathcal{N}(f^*, q^*)$), devemos tomar f^* e q^* tais que:

$$\begin{aligned} f^* &= -\psi(\tau_1^*) + \ln(\tau_2^*) \\ q^* &= \psi'(\tau_1^*). \end{aligned} \tag{7}$$

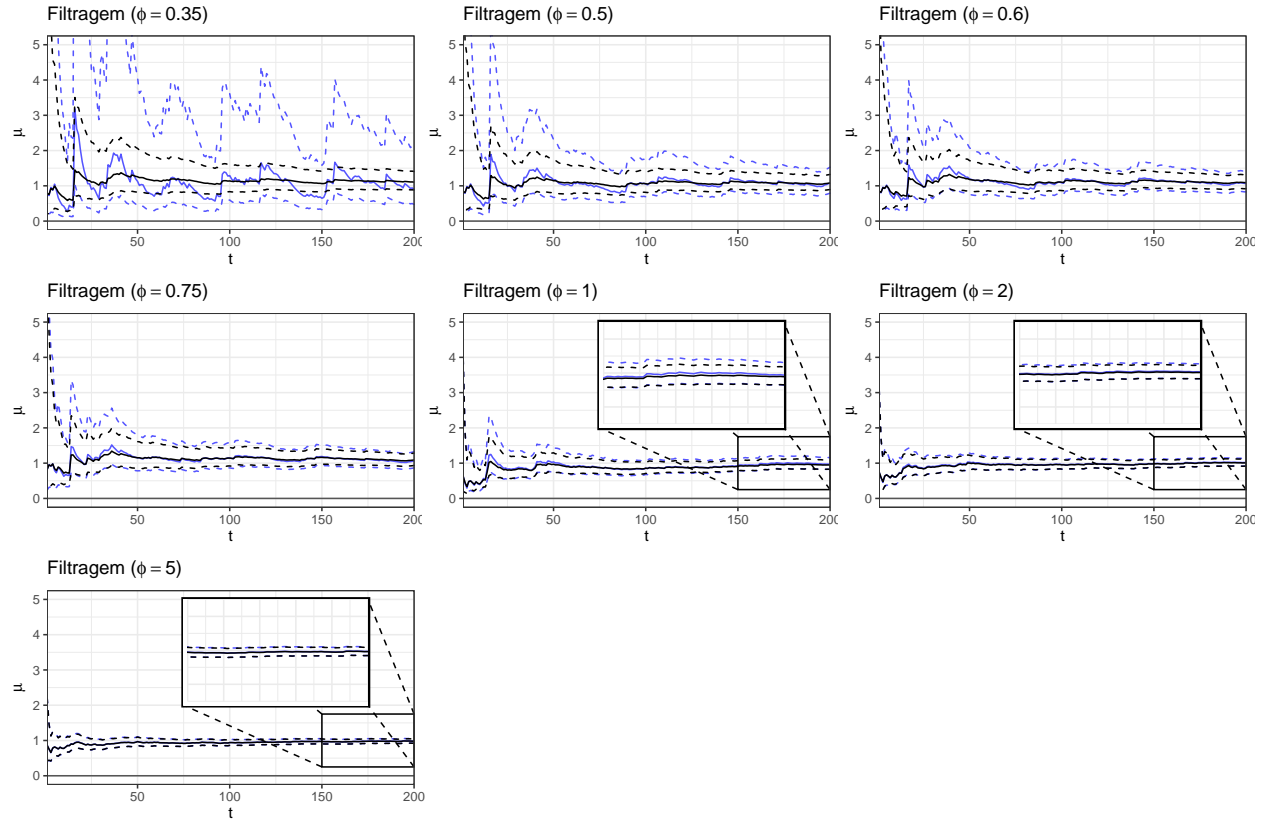
Observe que os sistemas obtidos em (6) e (7) são quase idênticos aos obtidos em (3) e (4), respectivamente, sendo que a única diferença é o sinal associada à f e f^* . De fato, isso se deve ao fato de que, se $X \sim \mathcal{IG}(\alpha, \beta)$, então $\frac{1}{X} \sim \mathcal{G}(\alpha, \beta)$, e se $X \sim \mathcal{LN}(\mu, \sigma^2)$, então $\frac{1}{X} \sim \mathcal{LN}(-\mu, \sigma^2)$.

Veja que a relação entre q e τ_1 permanece a mesma que no caso Poisson, sendo válidas todas as análises feitas à respeito (incluindo critérios para qualidade da aproximação).

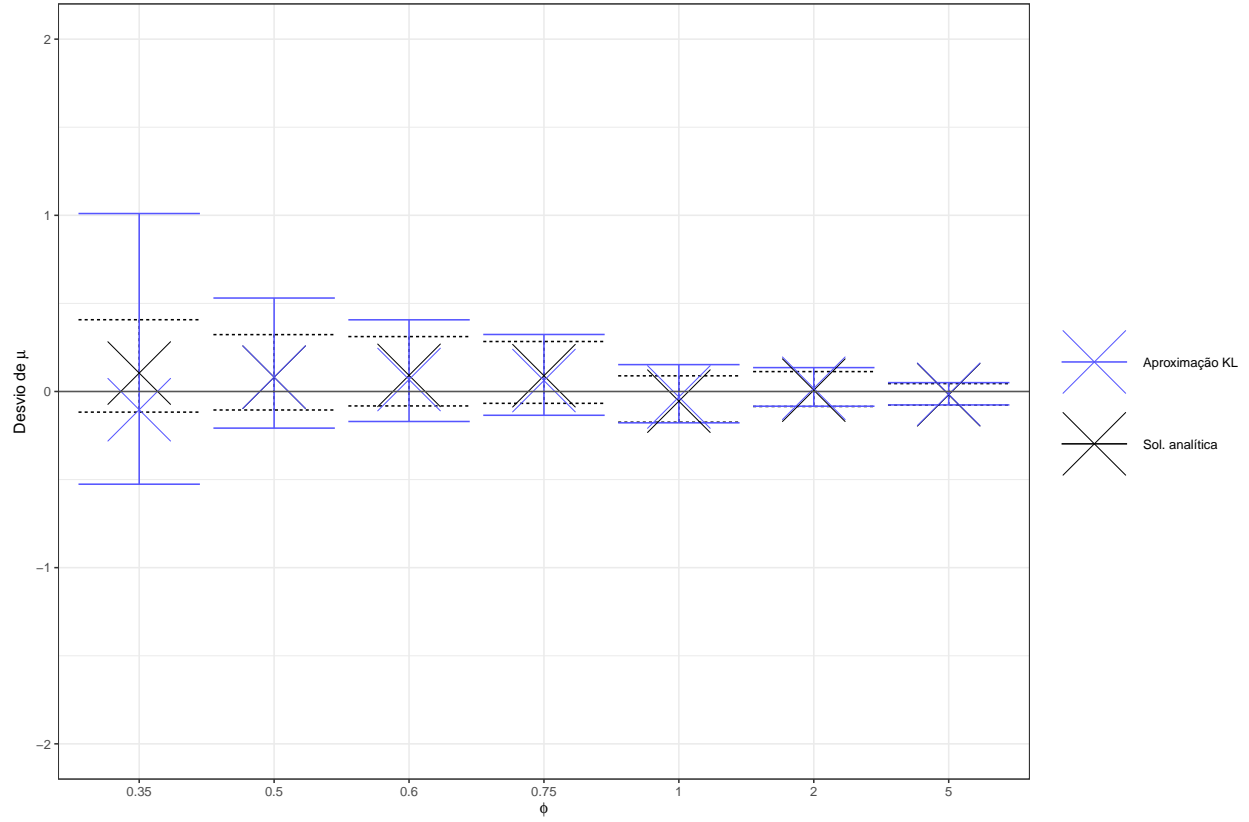
A grande diferença entre a qualidade da aproximação no caso Gamma e no caso Poisson está associada a atualização dos parâmetros, uma vez que $\tau_1^* = \tau_1 + \phi$ (ou seja, a qualidade da aproximação depende de ϕ , o parâmetro de forma o modelo observacional, e não da média dos dados observados). De fato, para **qualquer** que sejam as observações (desde que positivas), a qualidade do ajuste (no sentido de ser próximo ao caso analítico) será a mesma, sendo determinada exclusivamente por ϕ .

Nesta seção tentaremos encontrar o menor valor de ϕ tal que o ajuste é razoável. Naturalmente, para ϕ suficientemente grande, o ajuste será bom, porém, vale lembrar que, para ϕ grande, a distribuição Gamma converge à distribuição Normal, portanto, estamos especialmente interessados em testar valores pequenos de ϕ .

Adiante, exibimos a comparação entre o ajuste feito pela aproximação KL e o ajuste analítico para várias amostras i.i.d. com diversos valores de ϕ . Em todos os casos a média utilizada foi $\mu = 1$, sendo que o valor da média não é relevante para a qualidade do ajuste (i.e., a escala do dado não afeta a qualidade do ajuste).



As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.



Pelos gráficos acima podemos ver que o ajuste é bem satisfatório para $\phi \geq 0.75$ e parece bem decente para $0.5 \leq \phi < 0.75$.

De forma análoga ao caso Poisson, exibiremos ajustes em casos onde não há solução analítica (o valor de μ tem dinâmica temporal).

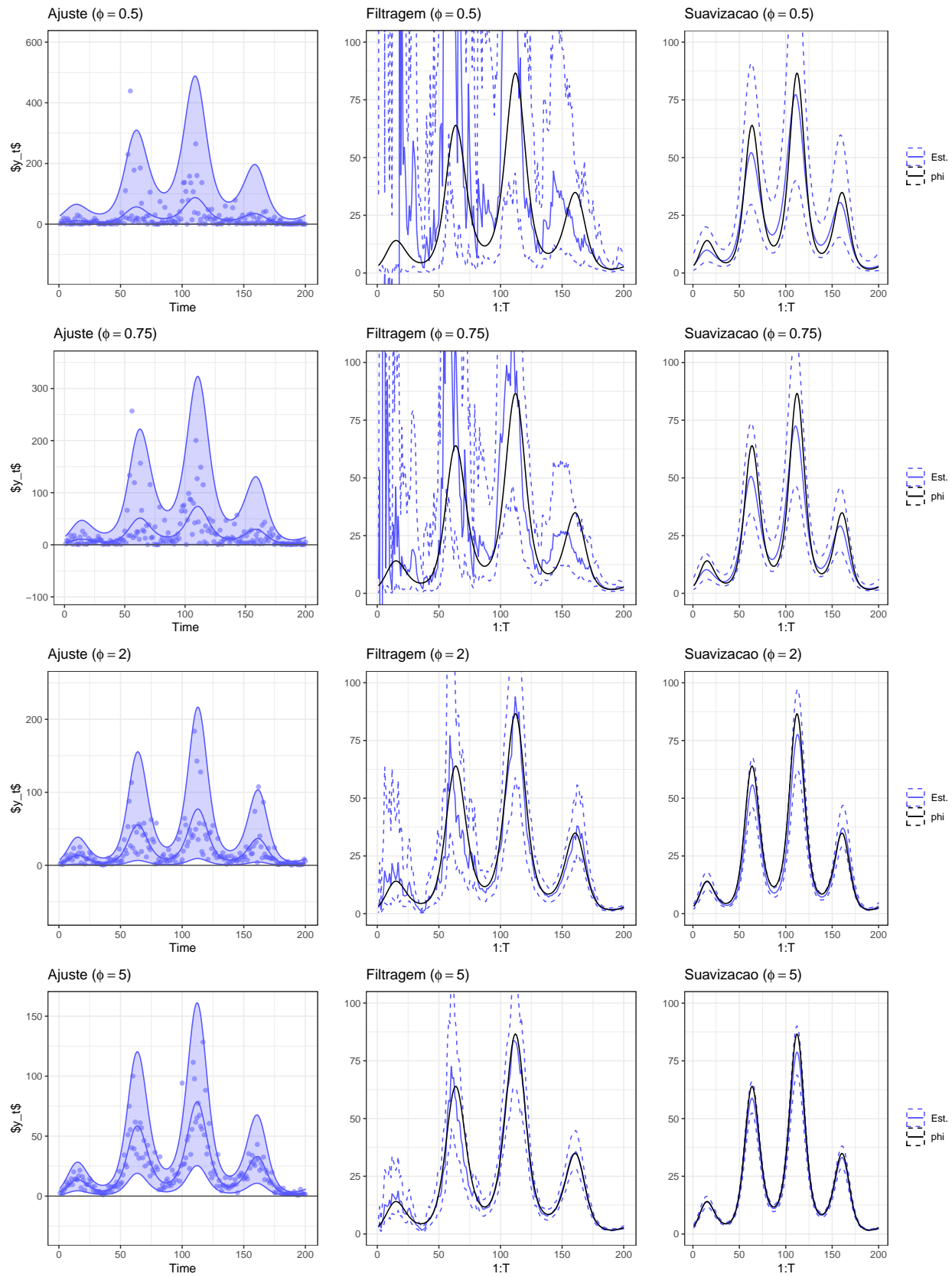
Amostraremos y_1, \dots, y_{200} com as distribuições a seguir:

$$y_i \sim \text{Poisson} \left(\phi, \frac{\phi}{\mu} \right),$$

com:

$$\mu = \frac{1}{10} \exp \left\{ \sin \left(\frac{i}{25} \pi \right) - \frac{i}{20} \left(\frac{i}{200} - 1 \right) + 1 \right\}.$$

Adiante exibimos o gráfico de μ ao longo do tempo junto com os valores amostrados e os ajustes para diversas escolhas de ϕ :



Qualidade da compatibilização de priori: Caso Multinomial

Nesta seção trataremos da qualidade do ajuste para o caso Multinomial. Para a análise deste caso, nos limitaremos ao caso Binomial, pois este caso já é o suficiente para determinar a qualidade da aproximação, desta forma trataremos do seguinte modelo:

$$\begin{aligned} y|\pi &\sim \text{Binom}(m, \pi) \\ \ln\left(\frac{\pi}{1-\pi}\right) &= \lambda \\ \lambda &\sim \mathcal{N}(f, q). \end{aligned}$$

Nesta notação, π é a probabilidade de sucesso, m é o número de ensaios e λ é o preditor linear.

Na abordagem KL, para obter a priori conjugada para π ($\pi \sim \mathcal{B}(\tau_1, \tau_2)$), devemos resolver o seguinte sistema:

$$\begin{aligned} \psi(\tau_1) - \psi(\tau_1 + \tau_2) &= f \\ \psi(\tau_2) - \psi(\tau_1 + \tau_2) &= -\ln(1 + e^f) + \frac{q}{2}e^f(1 + e^f), \end{aligned} \tag{8}$$

Para obter a solução para este sistema é necessário recorrer ao método Newton-Raphson (N-R). Durante a última reunião, mencionei que havia um problema no ajuste quando $\tau_1 + \tau_2$ é muito grande, de fato, parecia haver uma convergência da solução de modo a forçar $\tau_1 + \tau_2 < 400$, porém, durante a produção deste relatório, identifiquei que o problema era a tolerância da solução do N-R. Para uma estimação adequada dos parâmetros é necessário que tolerância seja bem baixa ($1e - 20$ funcionou bem), mas o *default* do algoritmo no *R* é relativamente alto ($1e - 6$).

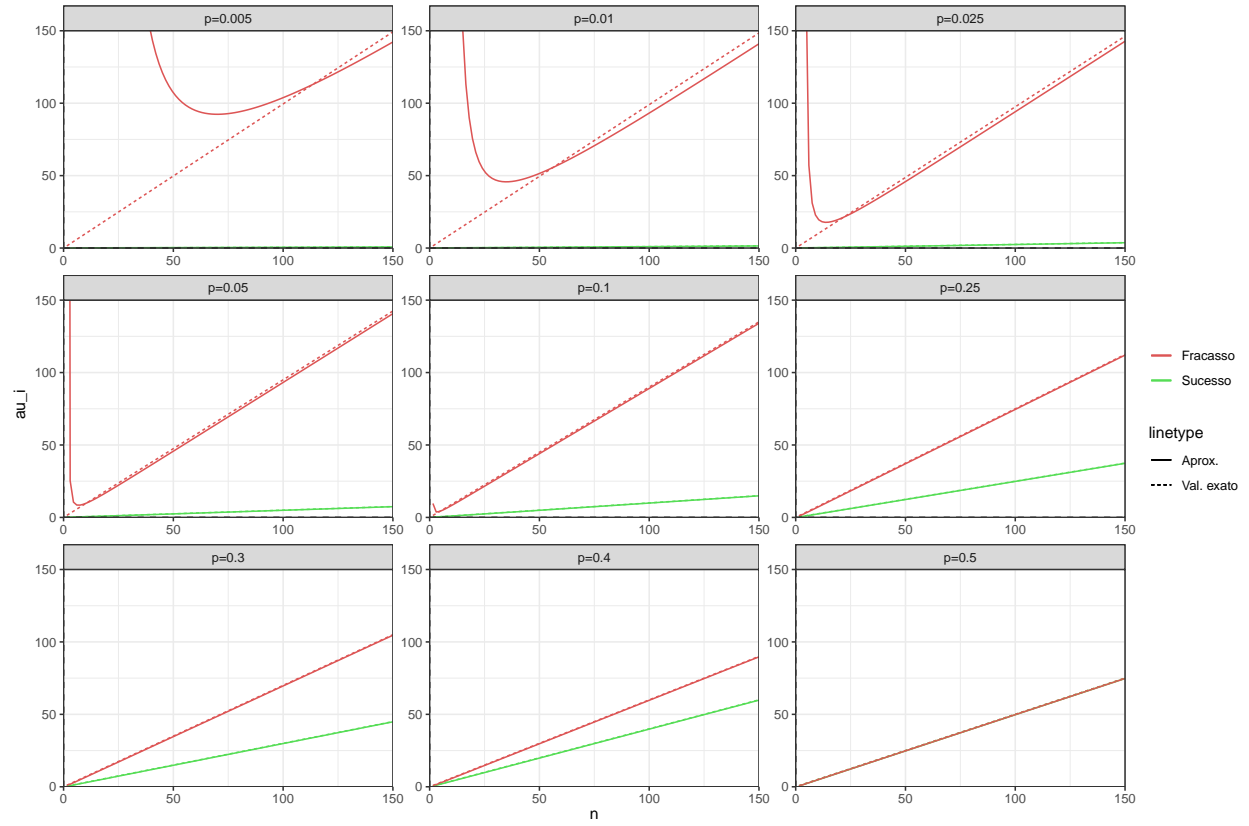
Obtidos τ_1 e τ_2 , definimos $\tau_1^* = \tau_1 + y$ e $\tau_2^* = \tau_2 + m - y$ de modo que $\pi|y \sim \mathcal{B}(\tau_1^*, \tau_2^*)$

Para obter a posteriori Normal para λ ($\lambda|y \sim \mathcal{N}(f^*, q^*)$), devemos tomar f^* e q^* tais que:

$$\begin{aligned} f^* &= \psi(\tau_1^*) - \psi(\tau_2^*) \\ q^* &= \psi'(\tau_1^*) + \psi'(\tau_2^*). \end{aligned} \tag{9}$$

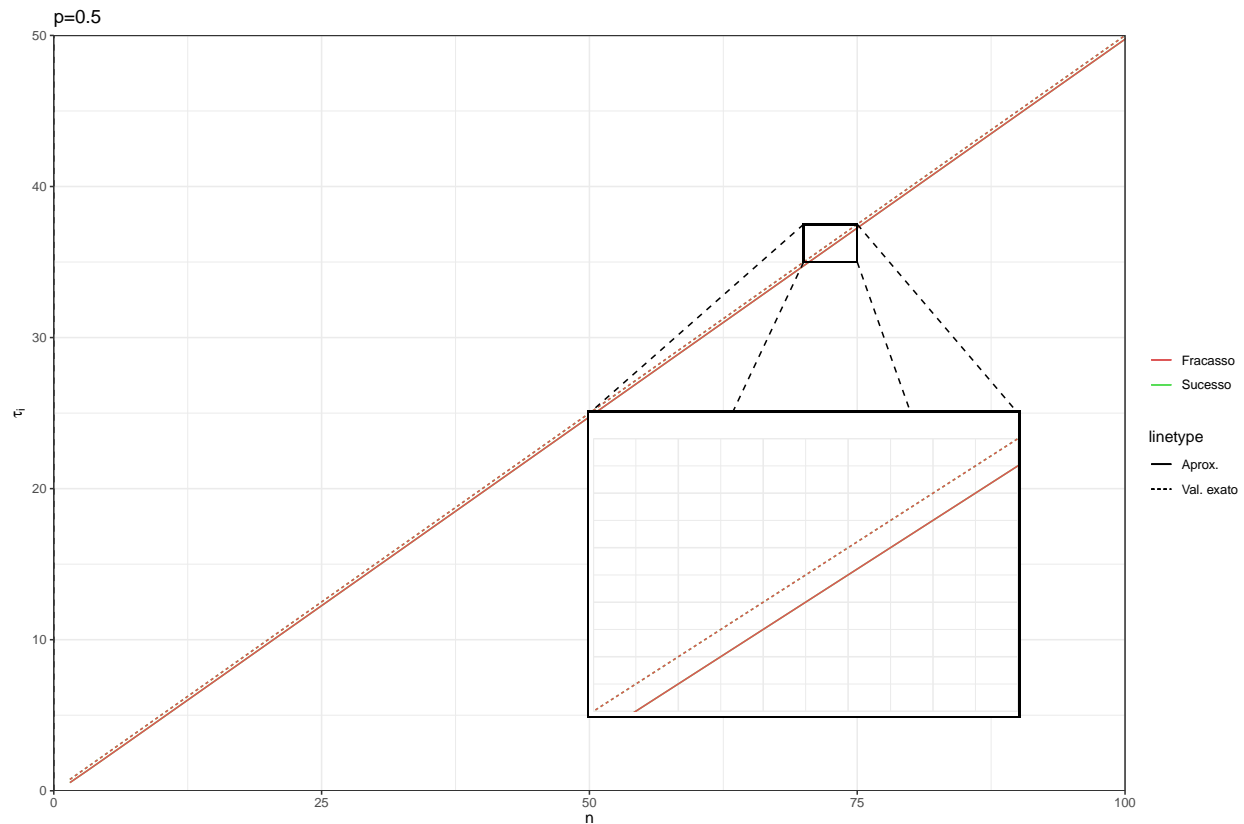
Para a análise da qualidade da aproximação, estamos interessados em descobrir quais os valores de m e π para os quais o ajuste fica satisfatório. Vale destacar que, por simetria, basta testar valores de $\pi \leq 0.5$.

O primeiro teste realizado é semelhante ao primeiro teste do caso Poisson: Começamos com uma distribuição Beta, aproximamos usando o método KL e então trazemos de volta para distribuição Beta usando o sistema de volta. Para cada teste, escolhemos $p \in (0, 0.5]$ e $n \in [0, 300]$, e realizamos o teste para uma Beta com parâmetros $\tau_1 = pn$ e $\tau_2 = (1 - p)n$. É importante destacar que, neste caso, p **não** é a probabilidade de sucesso (não confundir com π) e n **não** o número de ensaios (não confundir com m)! Estamos apenas reparametrizando a distribuição Beta para facilitar a análise, ainda não estamos gerando dados ou ajustando modelos. Podemos pensar em p como algo que representa a simetria ou assimetria da Beta e n como algo que representa a concentração da Beta. O intuito dos próximos gráficos é avaliar a qualidade da aproximação para distribuições Beta com precisão baixa e/ou bastante assimétricas.



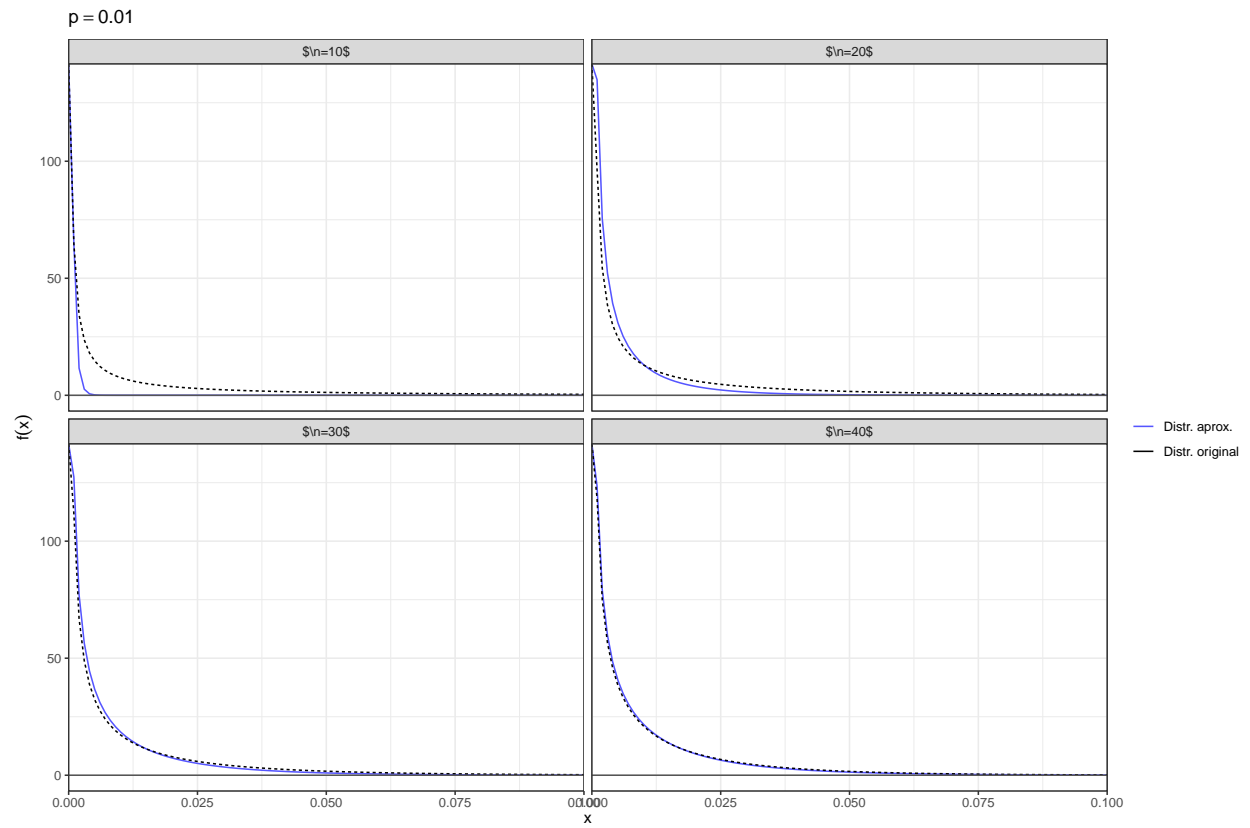
Veja que, para p razoavelmente grande (0.1) a aproximação parece conservar bem os parâmetro e, mesmo quando p é bem pequeno, basta que n seja adequadamente grande para que a aproximação fique boa.

Uma característica muito importante da aproximação KL é que há um certo deslocamento dos parâmetros após a conversão, como pode ser visto a seguir:

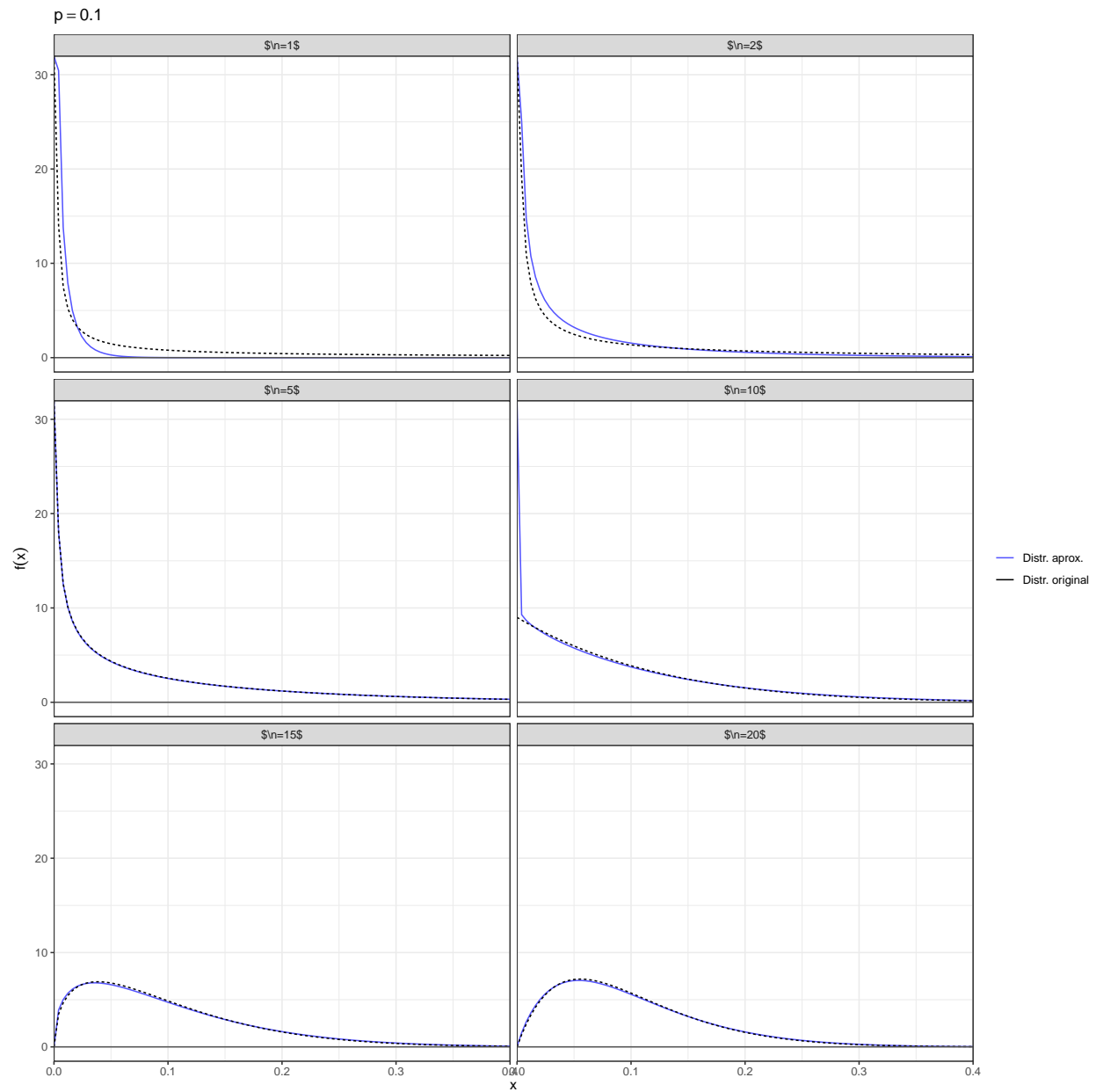


O descolamento da curva é em torno de 0.5 para baixo em relação ao valor exato, e é o mesmo para qualquer p desde que o n seja suficientemente grande (de fato esse deslocamento converge assintoticamente por cima para 0.5 conforme n aumenta). A princípio pode não parecer relevante esse deslocamento, afinal, 0.5 é pouco para n suficientemente grande, porém, veremos mais a frente que esse valor pode ser bem problemático para alguns valores de m , especificamente no caso Bernoulli onde $m = 1$.

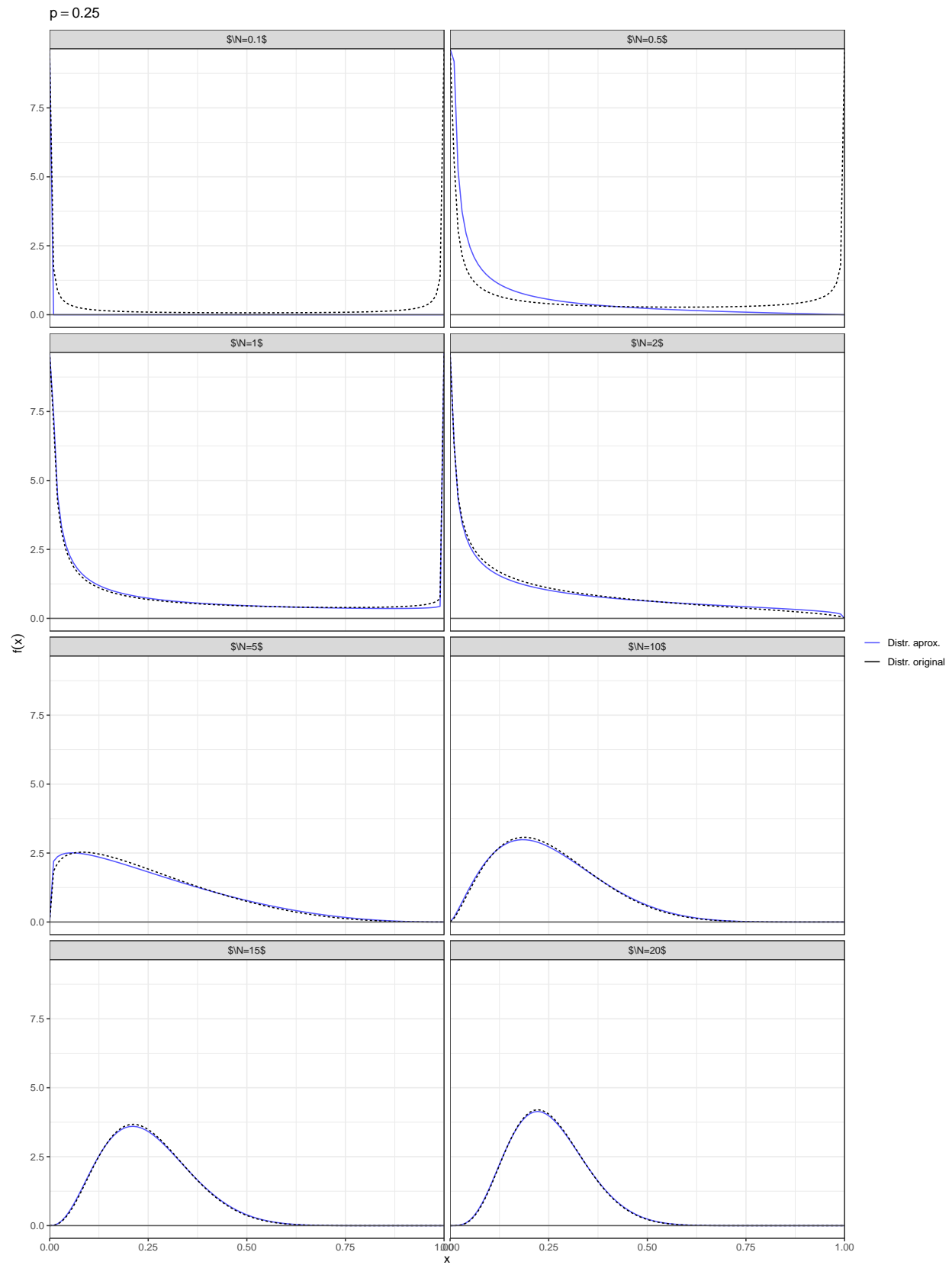
Avaliaremos agora a qualidade do ajuste olhando para a densidade da Beta obtida após a conversão:



podemos observar que, para $p = 0.01$, a aproximação começa a ficar boa por volta do $n = 20$, ou seja, com $\tau_1 = 0.2$.

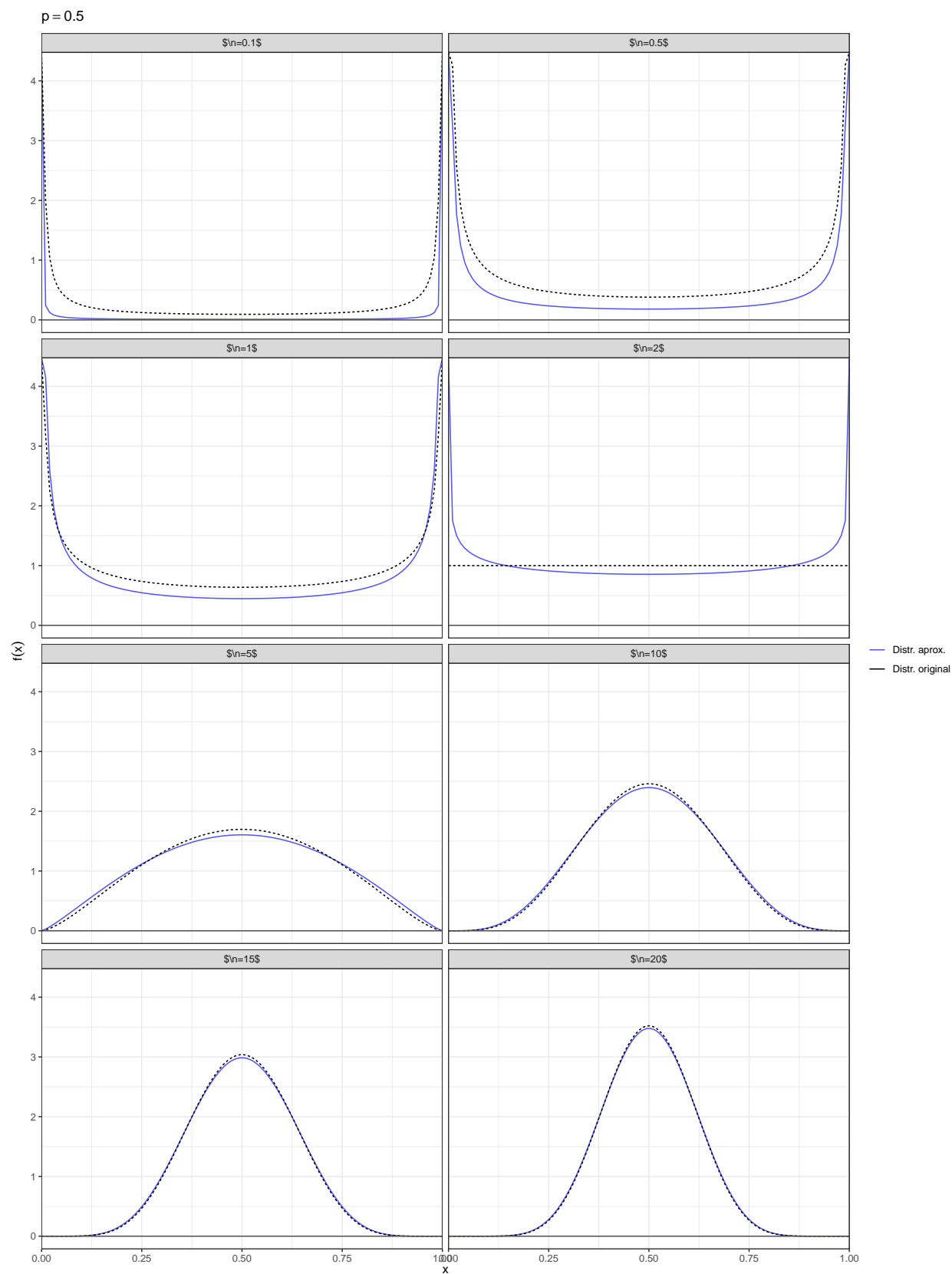


No gráfico acima podemos ver que, para $p = 0.1$, a aproximação começa a ficar razoável já para valores tão baixo quanto $n = 2$, sendo muito boa já para $n = 5$. Novamente, temos uma boa aproximação para valores de τ_1 tão baixos quanto 0.5 e 0.2.



Para $p = 0.25$ temos que a qualidade da aproximação fica boa já a partir de $n = 1$, novamente, temos que a

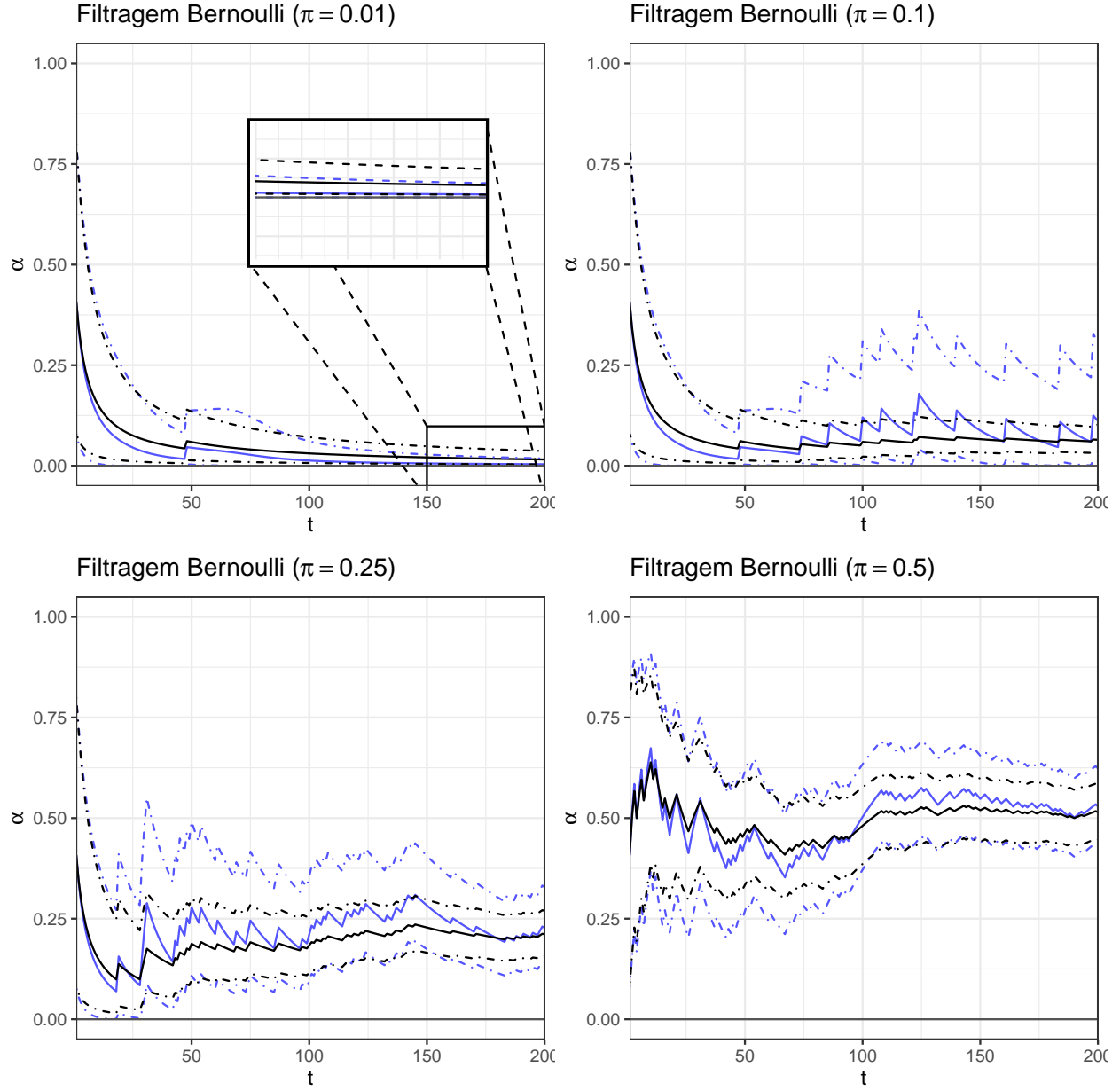
aproximação começa a ficar boa com $\tau_1 \approx 0.2$.



Pelo gráfico acima, podemos observar uma certa mudança de comportamento na qualidade da aproximação, que é razoável para $n = 0.5$ e $n = 1$, mas não parece tão boa para $n = 1$. Possivelmente a troca de concavidade da densidade pode ser responsável por isso, em especial, pode ser que a aproximação de segunda ordem usada para montar o sistema seja ruim próximo para certos valores de f onde há a mudança de concavidade na densidade. Mais análises são necessárias para verificar a extensão desse fenômeno e se ele de fato causa algum prejuízo. Dito isso, observe que, após uma iteração do processo de filtragem, $\tau_1 + \tau_2 > 1$, pois $m \geq 1$, logo, mesmo se esse caso for problemático, ele só pode causar problemas nas iterações iniciais.

Avaliaremos agora a qualidade do ajuste comparando os parâmetros estimados pela aproximação KL e os valores obtidos analiticamente para uma amostra i.i.d. de uma Binomial com os parâmetros adequados. Nesta etapa vamos avaliar a qualidade do ajuste para diferentes observações (anteriormente estávamos avaliando a qualidade da aproximação para diferentes priors). A ideia destes testes é encontrar quais os valores mínimos de m e π para os quais o modelo se ajusta bem (anteriormente estávamos avaliando quais os valores mínimos de n e p).

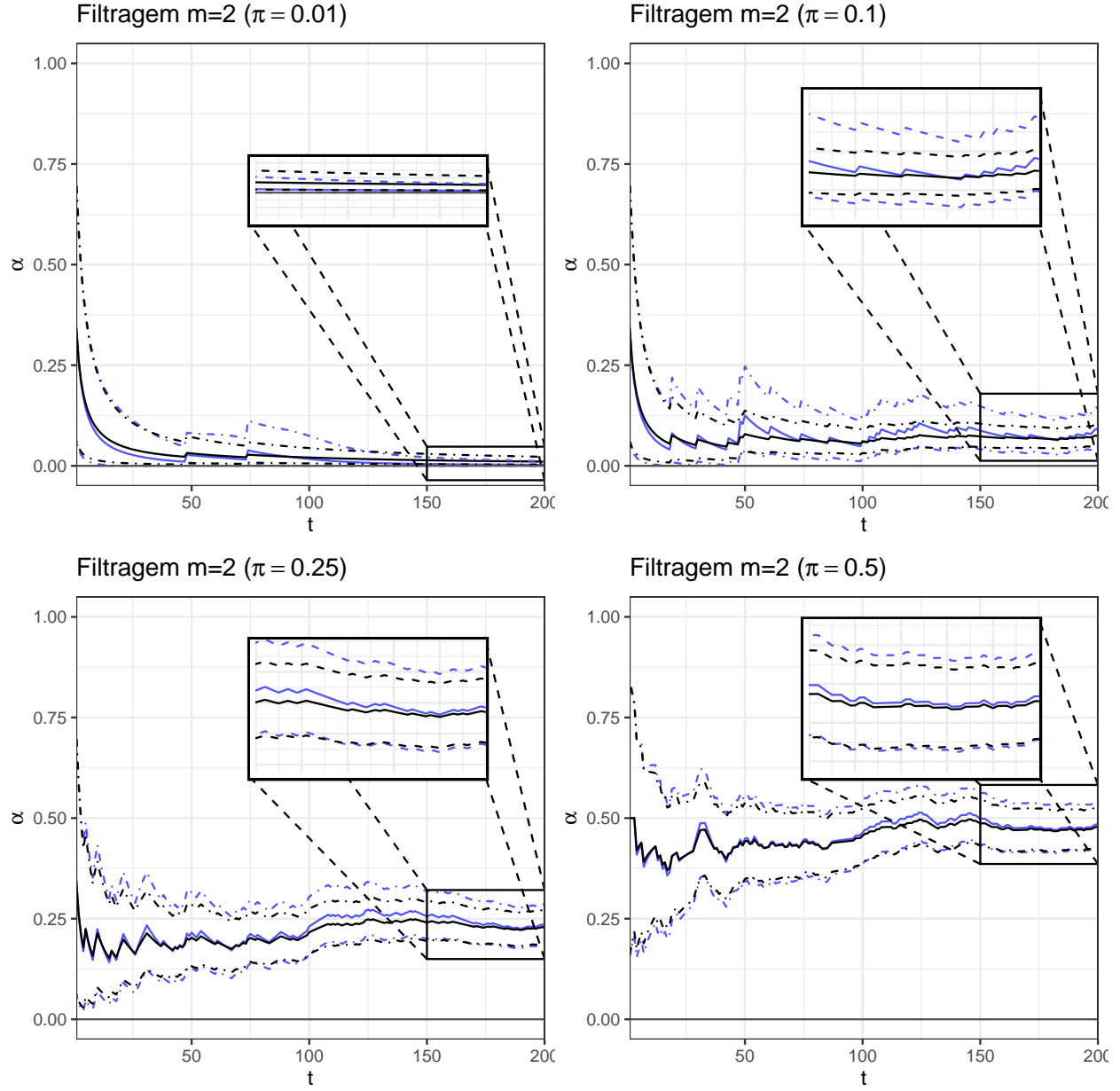
Caso $m = 1$ (Bernoulli):



As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Pelos gráficos acima podemos verificar que o ajuste para o caso Bernoulli fica significativamente diferente da solução analítica, mesmo para $\pi = 0.5$. A causa disso é o deslocamento observado nos parâmetros após a conversão da priori e da posteriori. Como mencionado anteriormente, quando a priori tem uma precisão minimamente razoável há um *shift* de 0.5 no valor $\tau_1 + \tau_2$ da distribuição conjugada. Veja que, no caso Bernoulli, $\tau_1 + \tau_2$ deveria aumentar em uma unidade a cada iteração, mas, por causa do *shift*, esse aumento acaba sendo de 0.5, assim, a solução pela aproximação KL tem o valor $\tau_1 + \tau_2$ igual a metade do ideal.

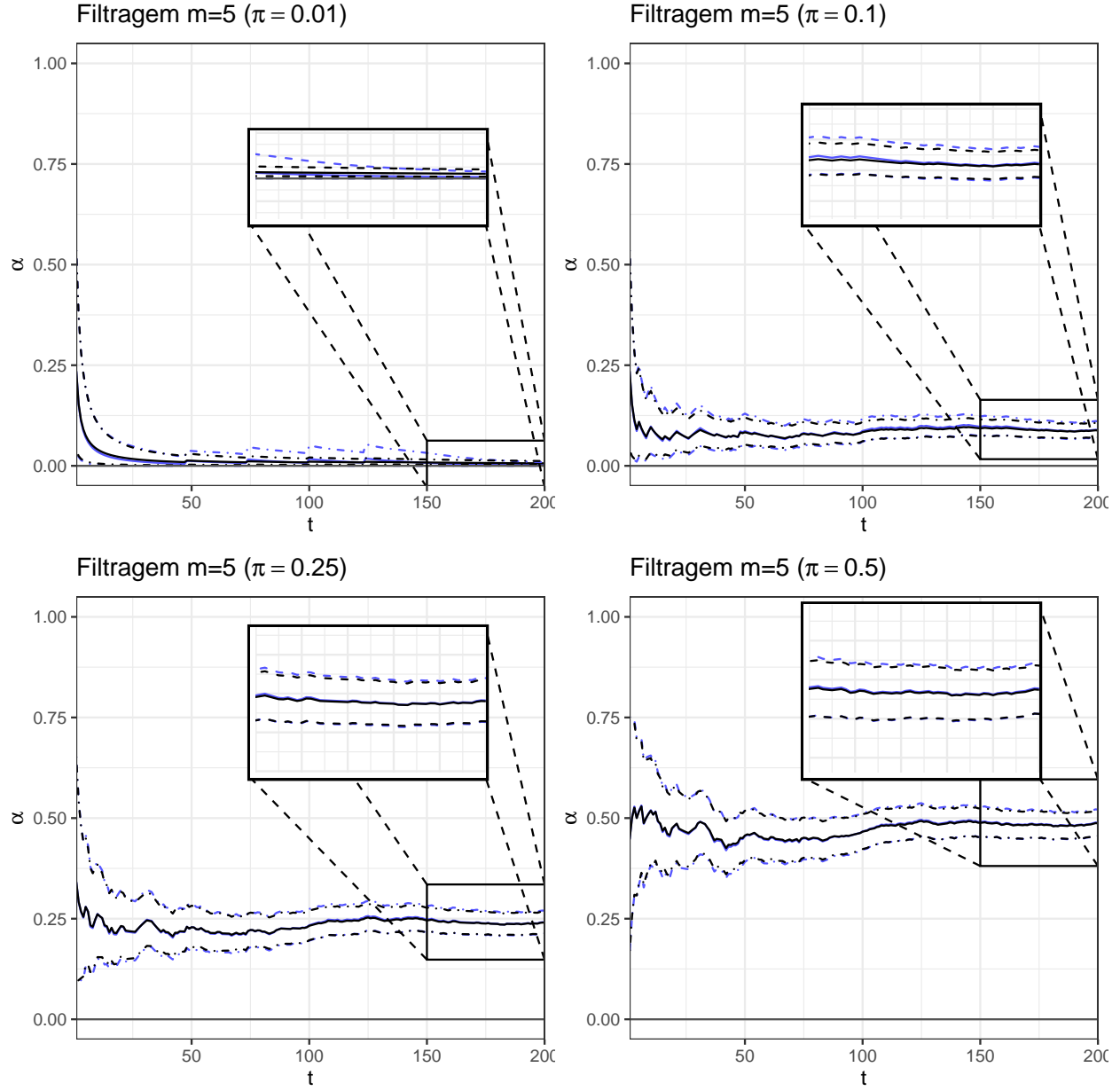
Caso $m = 2$:



As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

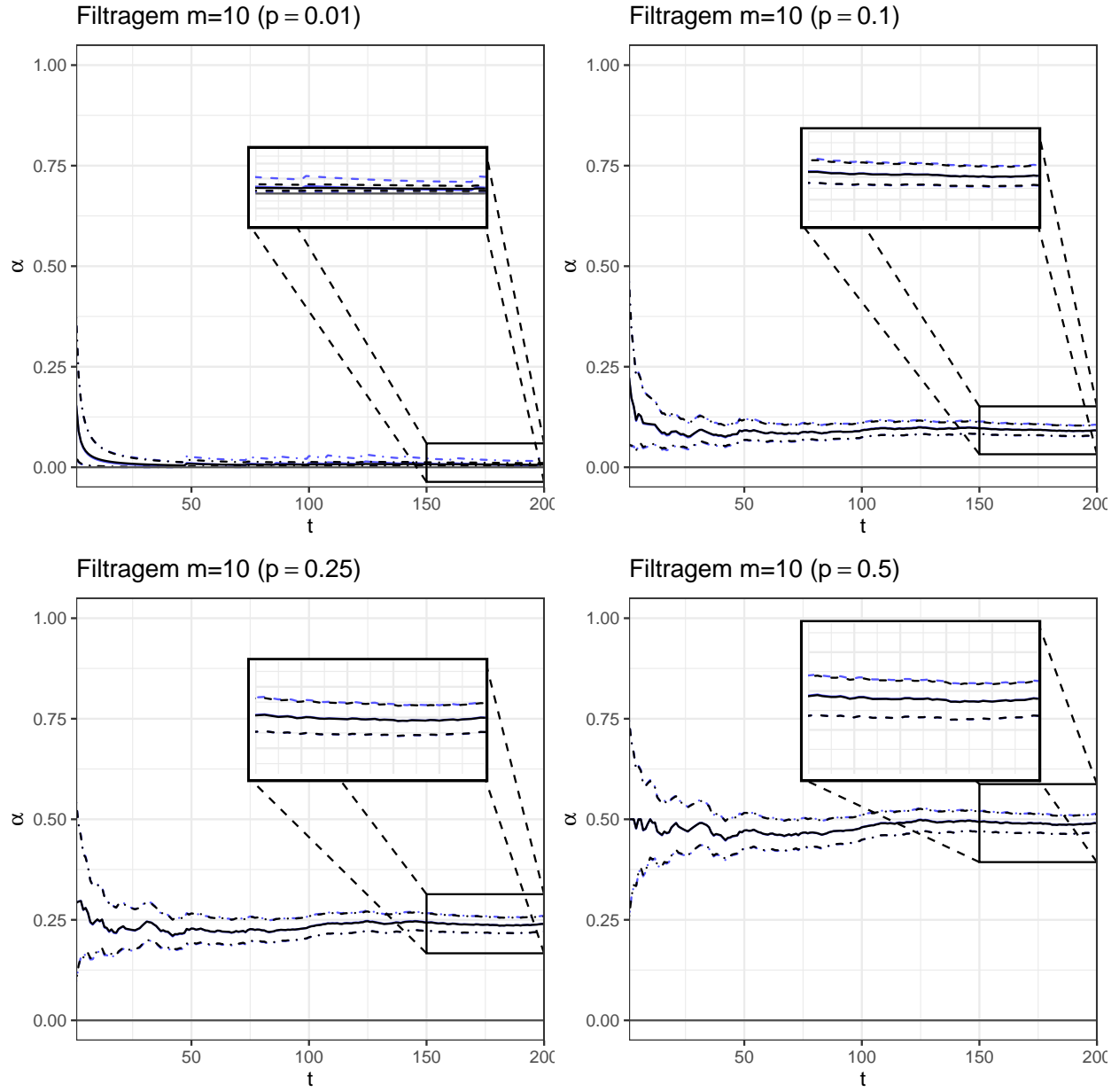
Podemos ver pelos gráficos acima que o *shift* dos parâmetros na conversão da priori tem um impacto muito menor quando $m = 2$, de fato, os ajustes parecem satisfatórios para $p \geq 0.1$. Vale destacar que, neste modelo, $m \in \mathbb{Z}_+$, logo a menos do caso Bernoulli, todos os outros ajustes serão pelo menos tão bons quanto este, uma vez que o caso Bernoulli é o único onde $m < 2$.

Caso $m = 5$:



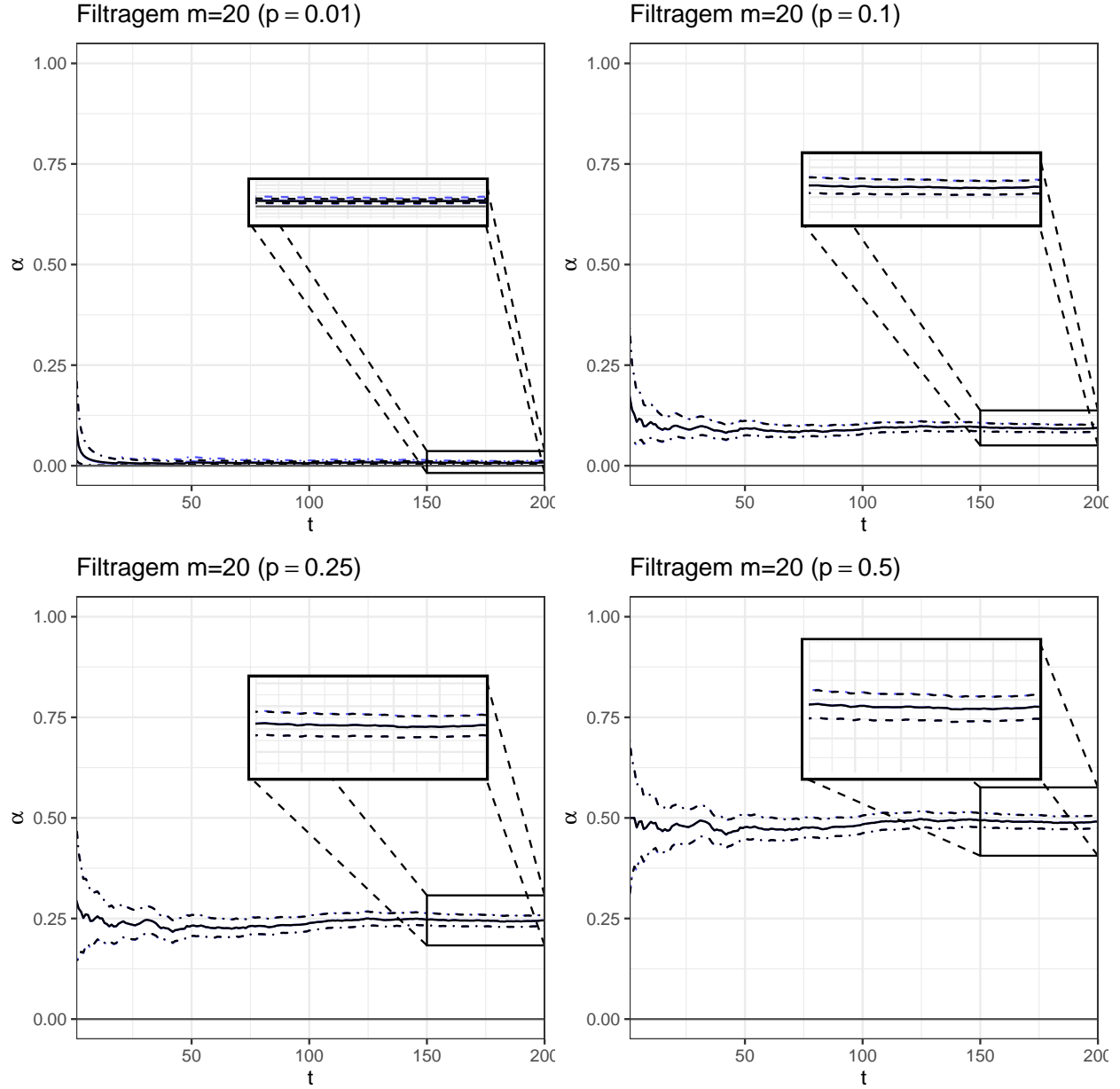
As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Caso $m = 10$:



As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Caso $m = 20$:



As curvas em preto representam o ajuste analítico em cada caso e a curva em azul representa o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

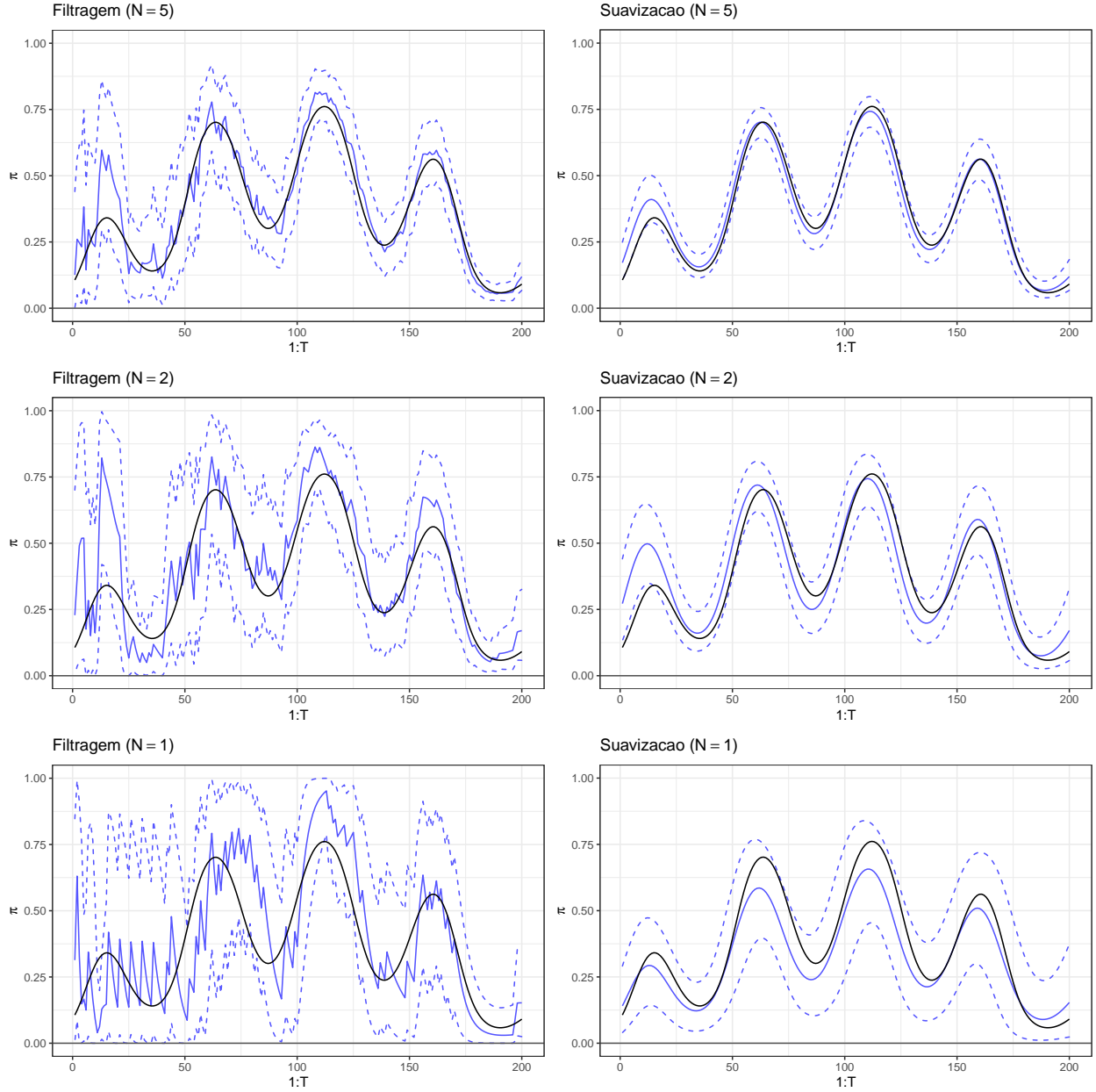
Neste momento, é relevante discutir a extensão deste caso para a Multinomial com mais de 2 categorias. Como observado nas análises da qualidade da aproximação, temos um resultado satisfatório desde que $\tau_1 \geq 0.2$. Esse resultado se estende para o caso Multinomial geral, não mostrei essa análise neste relatório para não ficar muito extenso, mas pude verificar que este é o caso, ademais, parece razoavelmente intuitivo dado a forma do sistema que um resultado assim aconteceria. Pelos gráficos acima, a menos do caso Bernoulli, o ajuste fica satisfatório desde que $\pi m \geq 0.2$ e esse resultado se estende para o caso de 3 ou mais categorias. De modo geral, uma euréstica para determinar se a aproximação das distribuições será satisfatória é avaliar se o valor esperado de ocorrências é maior ou igual a 0.2 em todas as categorias que se deseja modelar, se esta condição for satisfeita, então o ajuste ficará bom para todas as categorias. Caso uma ou mais categorias falhem em satisfazer essa condição, não há garantias de que o ajuste ficará bom. Nos casos em que observei, o não cumprimento da condição gerou superestimação da incerteza **apenas** nas categorias onde a condição

não vale, porém, mais análises são necessárias para verificar se este é sempre o caso.

Para finalizar esta seção, apresento o ajuste de 3 casos onde não há solução analítica. Em cada caso vamos usar um valor distinto de m (5,2,1) e a probabilidade de sucesso terá dinâmica temporal, de modo que:

$$\lambda_i = \frac{1}{10} \exp \left\{ \sin \left(\frac{i\pi}{25} \right) - \left(\frac{i}{20} \right) \left(\frac{i}{200} - 1 \right) \right\}$$

$$\pi_i = \frac{\lambda_i}{1 + \lambda_i}$$



As curvas em preto representam o valor real do parâmetro em cada caso e as curvas em azul representam o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Qualidade da compatibilização de priori: Caso Normal com variância desconhecida

Para finalizar este relatório, resta apenas lidar com o caso Normal com variância desconhecida. Este é, de fato, o caso com a pior qualidade de aproximação, ainda assim sob certas condições, o ajuste gerado me pareceu razoável. Entraremos nos detalhes sobre as condições para que o ajuste dos dados seja minimamente adequado, mas antes devemos apresentar o modelo utilizado e os sistemas a serem resolvidos afim de contextualizar o problema e fixar a notação:

$$\begin{aligned} y|\mu, \phi &\sim \text{Normal}(\mu, \phi) \\ \mu &= \lambda_1 \\ \ln(\phi) &= \lambda_2 \\ \lambda_1, \lambda_2 &\sim \mathcal{N}_2(f_1, f_2, q_1, q_2, \rho). \end{aligned}$$

Nesta notação, μ é a média dos dados, ϕ é a precisão, $\mathcal{N}_2(f_1, f_2, q_1, q_2, \rho)$ representa a distribuição Normal bivariada com médias f_1 e f_2 , variâncias q_1 e q_2 e correlação ρ (chamaremos também $q_{12} = \rho\sqrt{q_1 q_2}$) e $\lambda = (\lambda_1, \lambda_2)'$ é o preditor linear.

Na abordagem KL, para obter a priori conjugada para μ, ϕ ($\mu, \phi \sim \mathcal{NG}(\mu_0, c_0, \alpha, \beta)$), devemos resolver o seguinte sistema:

$$\mathbb{E}_{\mathcal{N}}[H_{\mathcal{NG}}] = \mathbb{E}_{\mathcal{NG}}[H_{\mathcal{NG}}], \quad (10)$$

onde $H_{\mathcal{NG}}$ é o vetor de estatísticas suficientes da distribuição Normal-Gamma, i.e.:

$$H_{\mathcal{NG}} = (\ln(\phi), \phi, \phi\mu^2, \phi\mu)'$$

Resolvendo o sistema (10) obtemos os seguintes parâmetros para a priori conjugada:

$$\begin{aligned} \mu_0 &= f_1 + q_{12} \\ c_0 &= \frac{1}{q_1} \exp\{-f_2 - q_2/2\} \\ \alpha &= \frac{1}{-3 + 3\sqrt{1 + \frac{2}{3}q_2}} \\ \beta &= \alpha \exp\{-f_2 - q_2/2\}, \end{aligned} \quad (11)$$

Obtidos os parâmetros da priori conjugada, podemos obter os parâmetros da posteriori conjugada a partir das seguintes equações:

$$\begin{aligned} \mu_0^* &= \frac{c_0\mu_0 + y}{c_0 + 1} \\ c_0^* &= c_0 + 1 \\ \alpha^* &= \alpha + 0.5 \\ \beta^* &= \beta + \frac{1}{2}c_0 \frac{(\mu_0 - y)^2}{c_0 + 1} \end{aligned} \quad (12)$$

Para obter a posteriori Normal para λ ($\lambda|y \sim \mathcal{N}_2(f_1^*, f_2^*, q_1^*, q_2^*, \rho^*)$), precisamos resolver o sistema a seguir:

$$\mathbb{E}_{\mathcal{N}}[H_{\mathcal{N}}] = \mathbb{E}_{\mathcal{NG}}[H_{\mathcal{N}}], \quad (13)$$

onde $H_{\mathcal{N}}$ é o vetor de estatísticas suficientes da distribuição Normal bivariada, i.e.:

$$H_{\mathcal{N}} = (\lambda_1, \lambda_2, \lambda_1^2, \lambda_2^2, \lambda_1 \lambda_2)'$$

Resolvendo o sistema (13) obtemos os seguintes parâmetros para a posteriori Normal::

$$\begin{aligned} f_1^* &= \mu_0^* \\ f_2^* &= \psi(\alpha^*) - \ln(\beta^*) \\ q_1^* &= \frac{\beta^*}{c_0^*(\alpha^* - 1)} \\ q_2^* &= \psi'(\alpha^*) \\ \rho^* &= 0 \end{aligned} \tag{14}$$

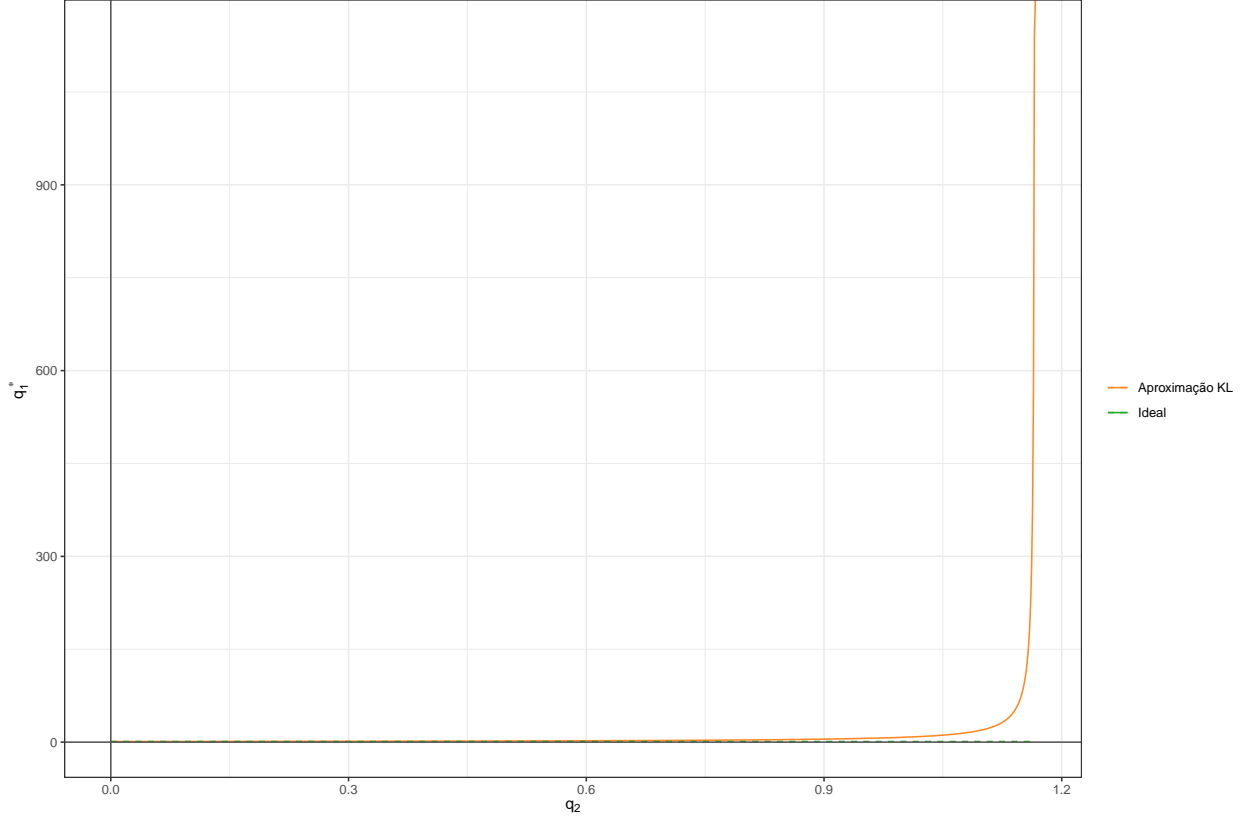
No restante desta seção estaremos interessados em encontrar as condições para que a aproximação KL seja satisfatória. Neste sentido, a primeira coisa a se observar é que a qualidade da aproximação **não** depende da escala ou locação dos dados. De fato, se tomamos $y^* = ay + b$, então a priori normal para o novo preditor linear será $\mathcal{N}_2(af_1 + b, f_2 - 2\ln(a), a^2 q_1, q_2, \rho)$, daí a priori conjugada será $\mathcal{NG}(a\mu_0 + b, c_0, \alpha, a^2 \beta)$, a posteriori conjugada será $\mathcal{NG}(a\mu_0^* + b, c_0^*, \alpha^*, a^2 \beta^*)$ e a posteriori normal será $\mathcal{N}_2(af_1^* + b, f_2^* - 2\ln(a), a^2 q_1^*, q_2^*, \rho^*)$, ou seja, mudar a escala dos dados e fazer o ajuste equivale a fazer o ajuste e mudar a escala dos dados. Com isto em mente, podemos fazer as análises sem nos preocupar com a escala ou locação dos dados.

De forma semelhante ao que foi feito no caso Poisson, a primeira coisa a ser verificada é se a aplicação da conversão de distribuições causa alguma degeneração nos parâmetros, especificamente, vamos avaliar o que ocorre com os parâmetros da Normal após a compatibilização quando não atualizamos a distribuição de μ, ϕ , isto é, se começamos com uma distribuição para λ_1, λ_2 , obtemos a distribuição conjugada para μ, ϕ e depois obtemos a distribuição λ_1^*, λ_2^* sem atualizar a distribuição de μ, ϕ , o quão diferente é a distribuição de λ_1, λ_2 e λ_1^*, λ_2^* ?

Para responder esta pergunta podemos observar que:

$$\begin{aligned} f_1^* &= \mu_0^* = \mu_0 = f_1 + q_{12} \\ f_2^* &= \psi(\alpha^*) - \ln(\beta^*) = \psi(\alpha) - \ln(\beta) = f_2 \\ q_1^* &= \frac{\beta^*}{c_0^*(\alpha^* - 1)} = \frac{\beta}{c_0(\alpha - 1)} = \frac{\alpha}{\alpha - 1} q_1 \\ &= \frac{\frac{1}{-3+3\sqrt{1+\frac{2}{3}q_2}}}{\frac{1}{-3+3\sqrt{1+\frac{2}{3}q_2}} - 1} q_1 \\ q_2^* &= \psi'(\alpha^*) = \psi'(\alpha) = \psi' \left(\frac{1}{-3+3\sqrt{1+\frac{2}{3}q_2}} \right) \\ \rho^* &= 0 \end{aligned}$$

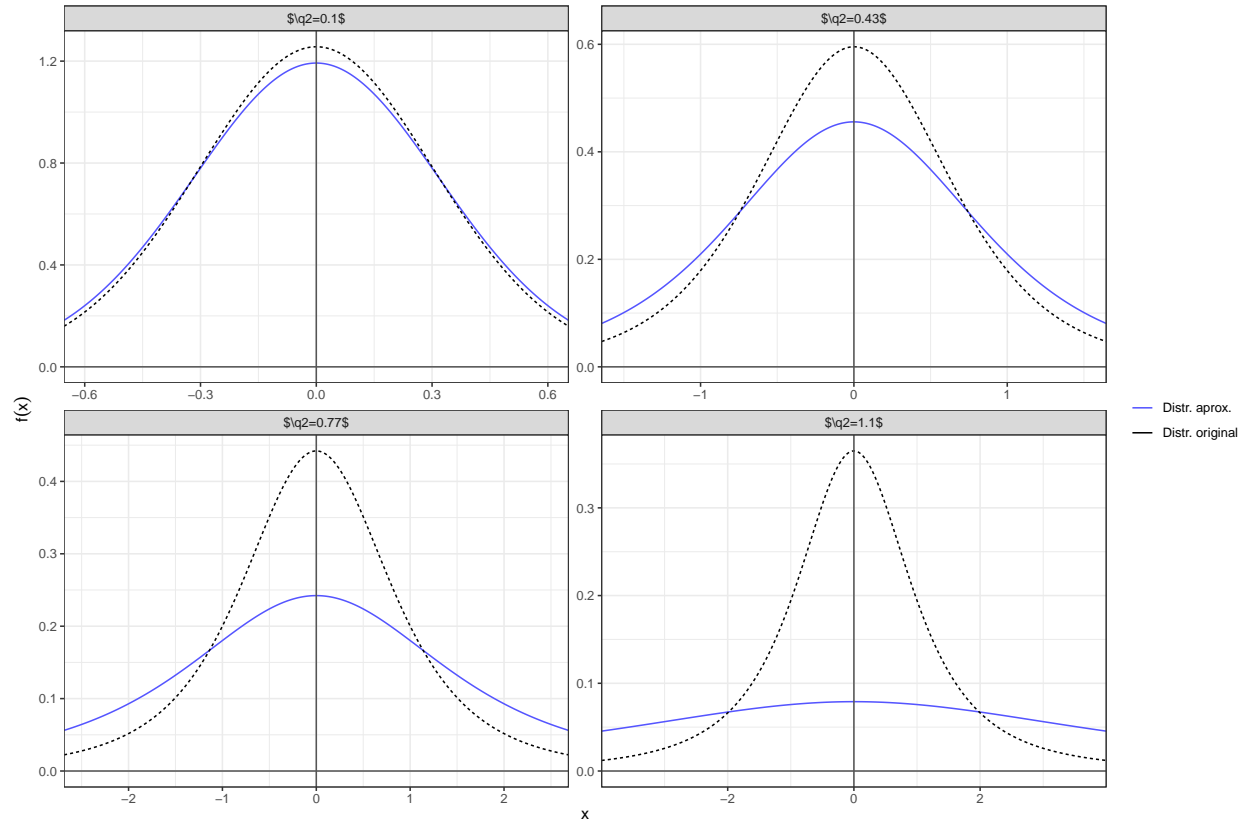
O comportamento de q_2^* é análogo ao comportamento descrito para a variância do preditor linear no caso Poisson. O efeito da compatibilização nas médias não é problemático, então resta avaliar a distorção em q_1 . Com mencionado anteriormente, a escala do dado não é relevante, portanto, vamos tratar do caso onde $q_1 = 1$ e avaliar o valor de q_1^* para vários valores de q_2 :



Observe que temos um problema muito grave quando q_2 se aproxima de $\frac{7}{6}$, pois neste caso α se aproxima de 1 e q_1^* tende a infinito, ademais, se $q_2 > \frac{7}{6}$, temos $\alpha < 1$ e $q_1^* < 0$! De fato, quando $\alpha < 1$, temos que $\mathbb{E}_{\mathcal{NG}}[\lambda_1^2] = \mathbb{E}_{\mathcal{NG}}[\mu^2] = +\infty$, pois a distribuição marginal de μ é t de Student com 2α graus de liberdade e, neste caso, μ não tem 2º momento finito, logo **não** é possível usar o Teorema da Projeção. De fato, pode ser verificado que **não existe** uma distribuição que minimiza a divergência KL, pois a divergência será uma função decrescente de q_2 . Há uma forma alternativa de lidar com esta situação que resolve todos estes problemas, porém, antes de apresentar esta proposta, continuaremos com a análise da qualidade do ajuste para os casos onde $\alpha > 1$.

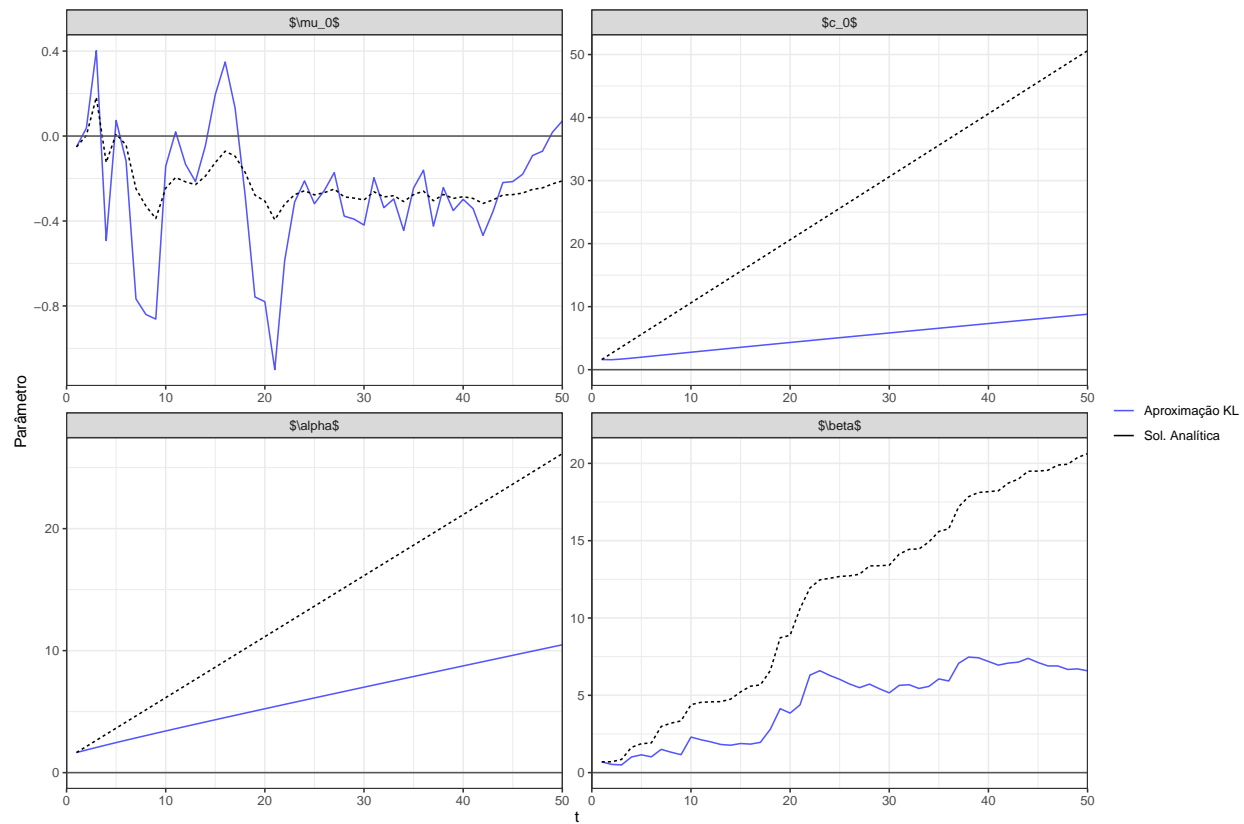
Como q_2 é a variância de λ_2 , é simples definir casos onde $q_2 < \frac{7}{6}$, de fato, o valor de q_2 dependerá exclusivamente da estrutura do modelo. Caso a estrutura inclua regressoras para ϕ , não é possível garantir que q_2 não exceda $\frac{7}{6}$ em algum momento, uma vez que q_2 dependerá do valor observado para a regressora em questão. Contudo, no caso onde não estamos usando regressoras, fica relativamente fácil controlar o valor de q_2 , porém, infelizmente, isso significa que não podemos usar prioris muito vagas e/ou estruturas muito complexas para ϕ . Dito isso, conseguimos fazer um ajuste razoável para casos onde ϕ não apresenta tendência de crescimento, mesmo incluindo uma estrutura sazonal para ϕ . Nos casos onde ϕ apresenta tendência de crescimento, é necessário que a priori para o nível do modelo tenha variância baixa. Vale destacar que, em qualquer caso que seja, **não** há restrições sobre a estrutura associada a μ , desta forma, é possível modelar problemas onde o comportamento da média é bem complexo, desde que a variância dos dados **não** precise de muita estrutura.

Adiante, comparamos a distribuição do parâmetro μ antes e depois da compatibilização das distribuições para vários valores iniciais de q_2 (valores válidos). No teste a seguir, começamos com a distribuição \mathcal{NG} (original), obtemos a distribuição Normal associada e depois obtemos a distribuição \mathcal{NG} (aproximada) correspondente. Vale lembrar que, usando a distribuição \mathcal{NG} para μ e ϕ , temos que a distribuição marginal de μ é $t\left(2\alpha, \mu_0, \frac{\beta}{c_0\alpha}\right)$.

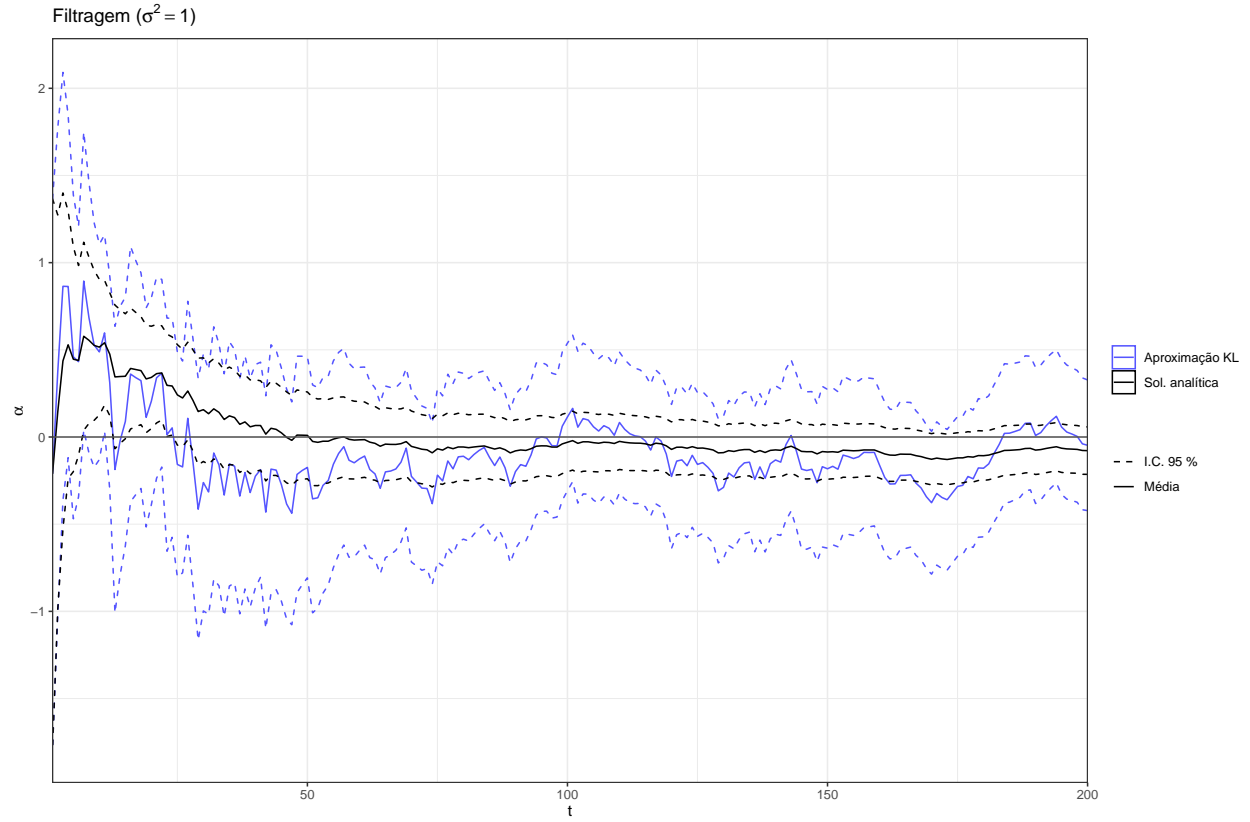


Veja que, quando q_2 é próximo de $\frac{7}{6}$ a aproximação é muito ruim. Tanto a distribuição aproximada quanto a original são t , mas durante a conversão o valor de q_1 é drasticamente inflado, o que faz com que o c_0 da aproximação seja muito menor do que o c_0 original.

Adiante comparamos a estimação dos parâmetros usando a abordagem KL com a estimação analítica para uma amostra i.i.d. de uma Normal Padrão:



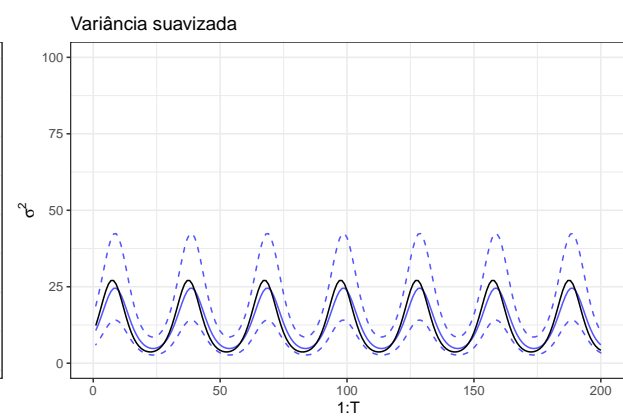
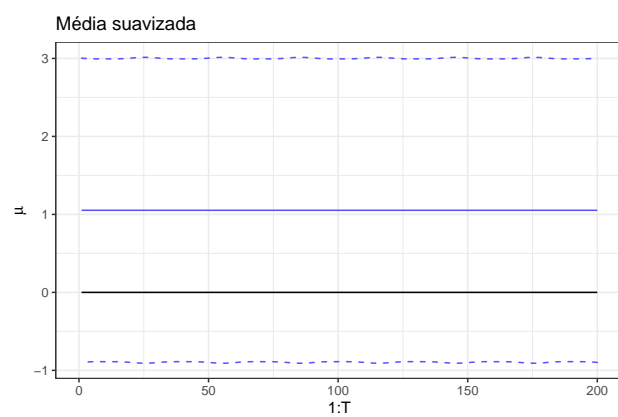
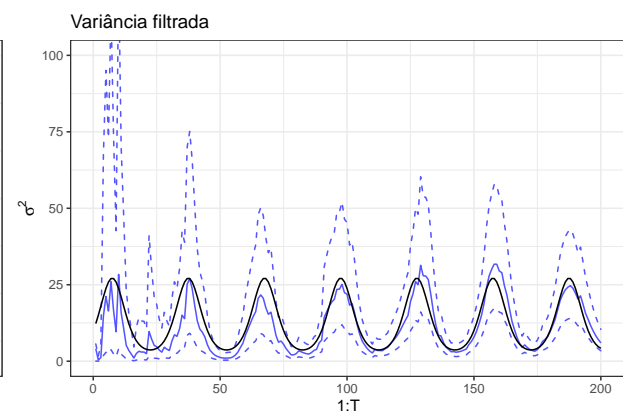
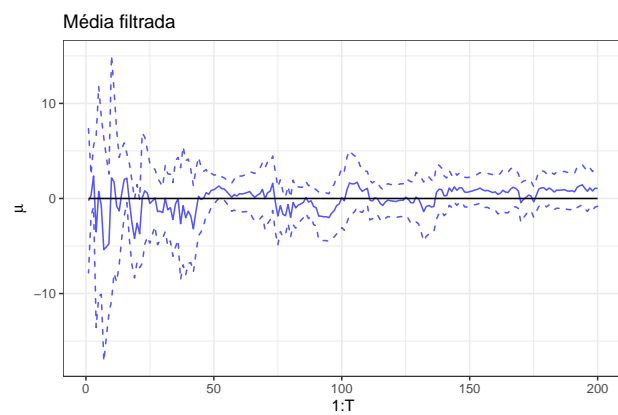
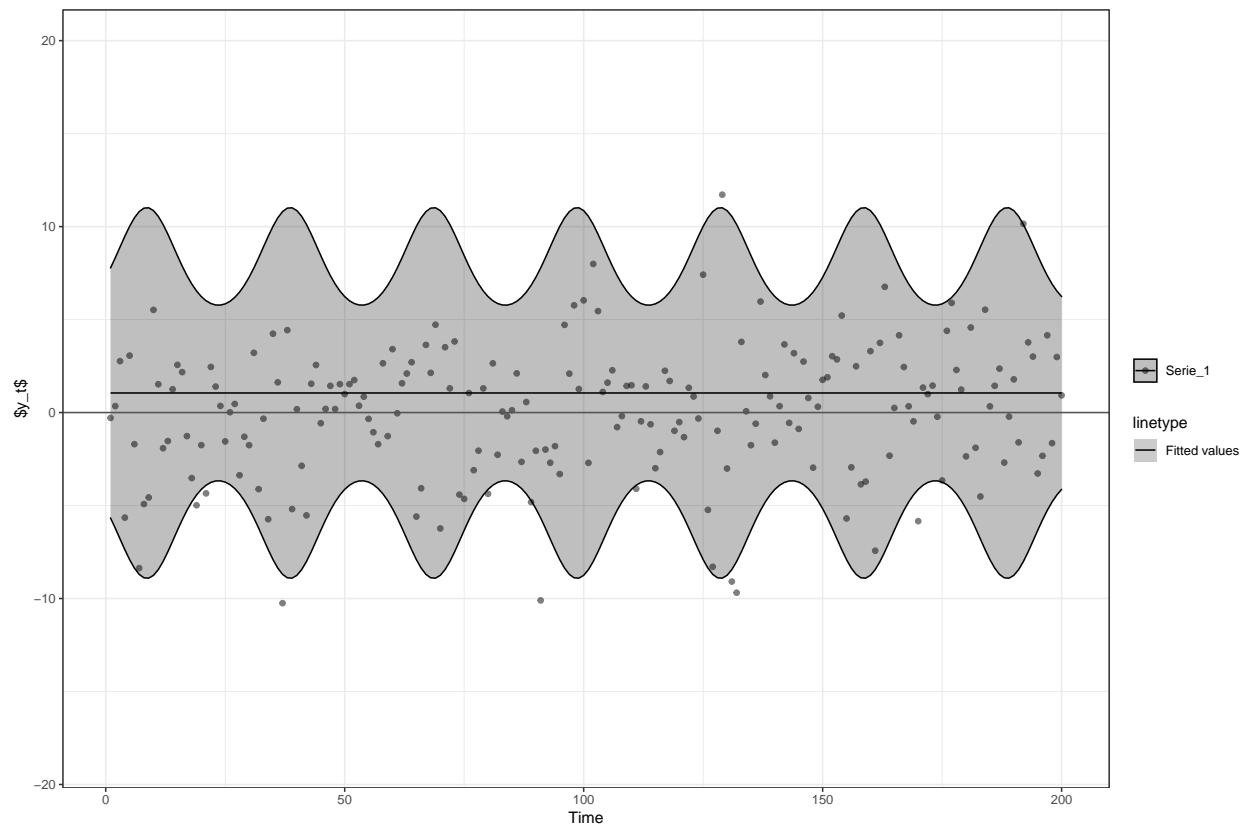
Olhando mais de perto para o parâmetro μ_0 , temos a seguinte estimativa intervalar e pontual para os dois ajustes para uma nova amostra de tamanho maior que a anterior:



Veja que o ajuste dos dados não fica tão ruim (dado os valores no espaço paramétrico, é de se imaginar que o ajuste fosse pior), gerando um resultado que me parece aceitável.

Por último, vamos analisar um caso onde não há solução analítica para verificar se a abordagem KL é capaz de capturar a dinâmica temporal dos dados, em especial da variância observacional. Para tal, geramos uma amostra y_1, \dots, y_{200} , tal que $y_i \sim \mathcal{N}(0, S_i)$ com:

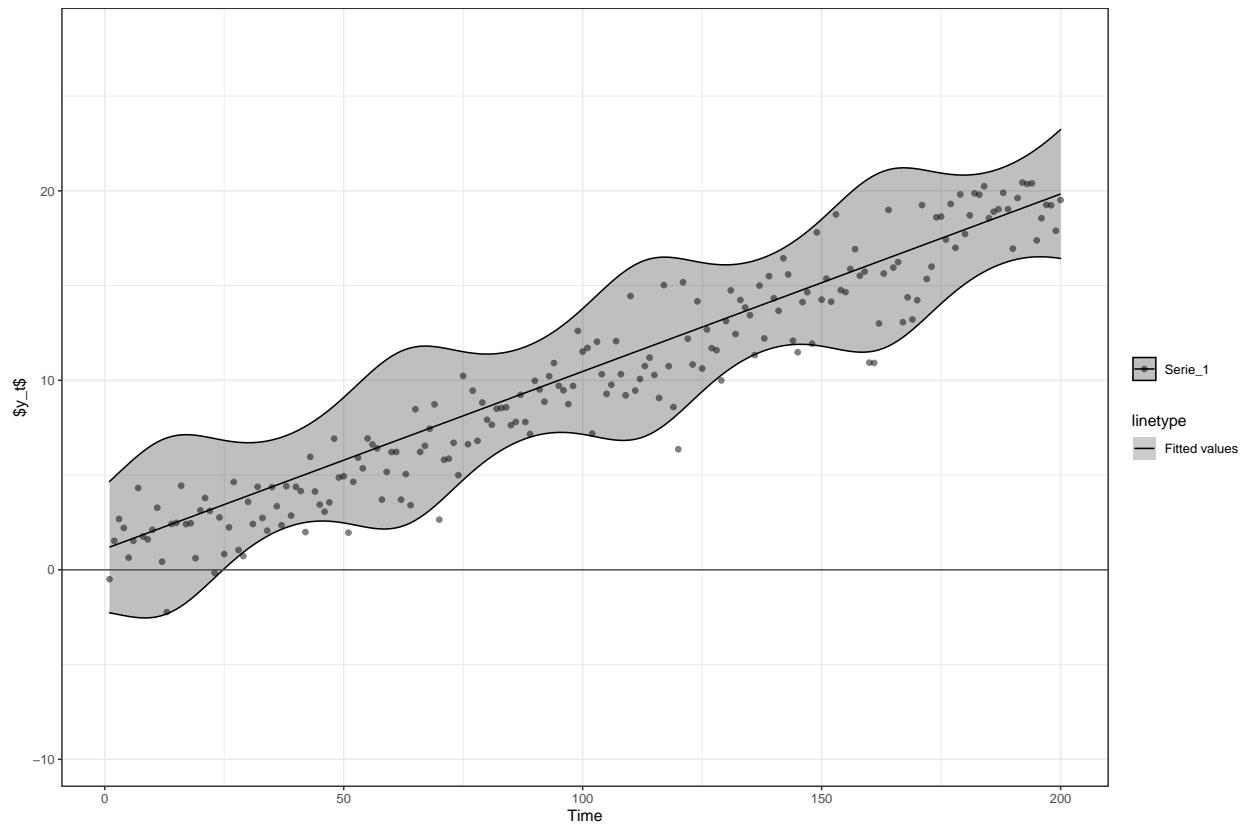
$$S_i = 10 \exp \left\{ \sin \left(\frac{i\pi}{15} \right) \right\}$$

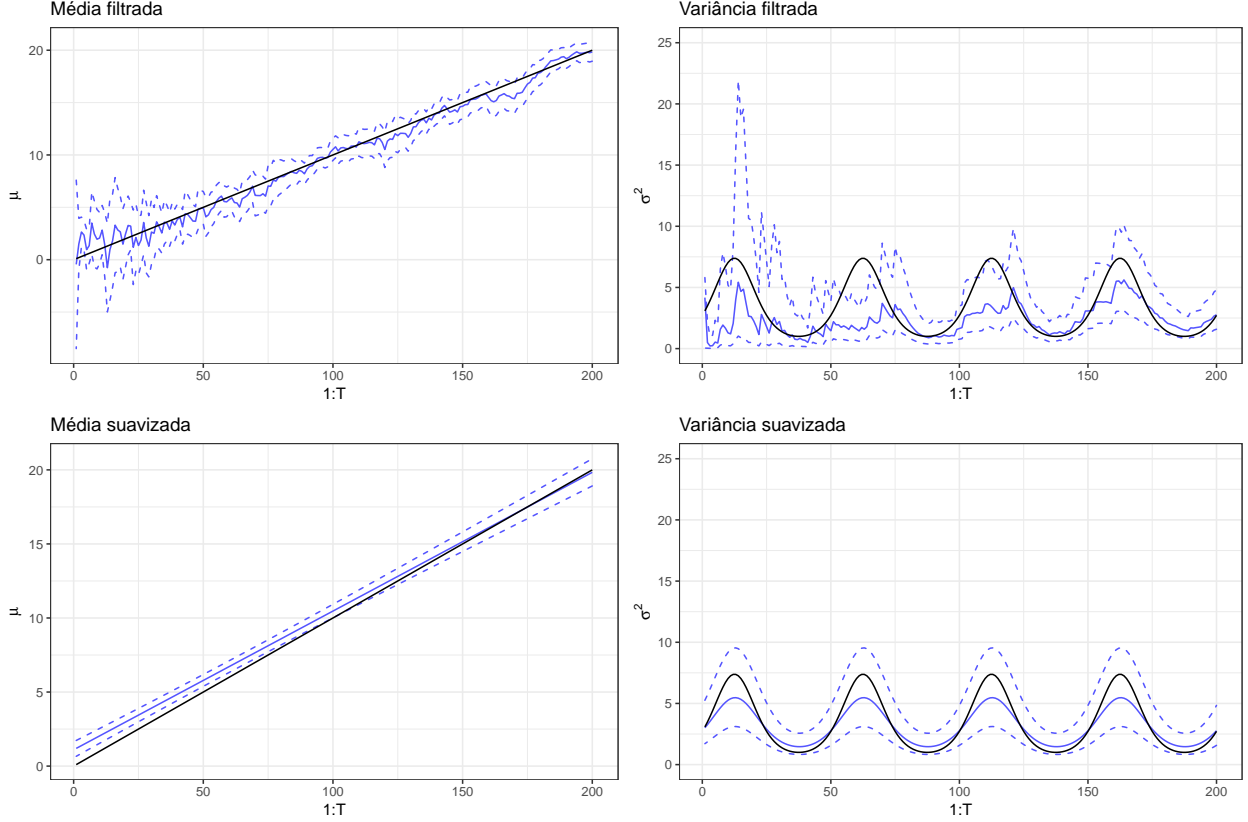


As curvas em preto representam o valor real do parâmetro em cada tempo e as curvas em azul representam o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Veja no resultado acima que o modelo consegue capturar bem o comportamento da variância e a dispersão dos dados, provavelmente não tão bem quanto se um método de MCMC fosse utilizado, mas ainda assim o ajuste parece satisfatório.

Adiante mostraremos o ajuste para uma situação parecida, mas agora com um modelo de crescimento linear para a média:





As curvas em preto representam o valor real do parâmetro em cada tempo e as curvas em azul representam o ajuste com a abordagem KL, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Novamente, as estimativas parecem bem descentes e o modelo parece se adequar bem aos dados observados. De modo geral, acredito que este ajuste seja válido, desde que a modelagem da variância não seja demasiadamente complexa. Em termos prático, uma vez que as condições necessárias para um ajuste aceitável (aceitável na minha perspectiva, claro) estão estabelecidas, é fácil implementar no pacote GDLM restrições sob a modelagem da variância, de modo a impedir que um usuário caia em casos inválidos. De todo modo, adiante apresentarei uma proposta que contorna o problema encontrado na qualidade do ajuste, oferecendo um ajuste mais próximo da solução analítica (nos casos onde podemos calculá-la), válido para quaisquer q_2 e α e sem restrições quanto a complexidade da estrutura do modelo, permitindo inclusive o uso de regressoras na modelagem de ϕ .

O problema que estamos encontrando na compatibilização das distribuições de λ_1, λ_2 e μ, ϕ tem sua origem no fato de que, quando $\alpha \leq 1$, não existem $f_1^*, f_2^*, q_1^*, q_2^*, \rho^*$ que minimize a divergência KL na volta (isto é, quando saímos da \mathcal{NG} e vamos para a normal bivariada), de fato, dados $f_1^*, f_2^*, q_1^*, q_2^*, \rho^*$, podemos sempre diminuir a divergência KL ao aumentar o valor de q_1^* e isso se deve ao fato de que, quando $\alpha \leq 1$, μ não tem 2º momento finito.

Pensando um pouco na proposta do artigo, a ideia é que, para aproximar uma distribuição p pertencente à família exponencial por outra distribuição em um espaço Ω de distribuições pertencentes à família exponencial, devemos tomar $q \in \Omega$ tal que q minimize $KL(p||q)$, sendo que a escolha de Ω é, em certo sentido, arbitrária, sendo usado um conjunto de distribuições com propriedades convenientes. A minha proposta para a resolução do problema é restringir o conjunto Ω para um conjunto onde a divergência KL possa ser minimizada, i.e., onde existe $q \in \Omega$ que minimize a divergência KL.

No contexto original do caso Normal com variância desconhecida, para irmos da posteriori conjugada para a posteriori normal, usamos (desculpe a notação) $\Omega = \{\mathcal{N}_2(f_1, f_2, q_1, q_2, \rho) : f_1, f_2 \in \mathbb{R}, q_1, q_2 \in (0, +\infty), \rho \in$

$(-1, 1)$. Minha proposta é restringir Ω ao conjunto das Normais bivariadas com $q_1 = c$, de modo que devemos escolher $q \in \Omega' = \{\mathcal{N}_2(f_1, f_2, q_1, q_2, \rho) \in \Omega : q_1 = c\}$. Primeiro, afirmo que neste conjunto existe q minimal e que podemos usar o Teorema da Projeção para encontrar q , de fato, basta observar que, sob esta restrição, o vetor H_q de estatísticas suficientes será dado por:

$$H_q = (\lambda_1, \lambda_2, \lambda_2^2, \rho),$$

sendo que todas estas variáveis tem valor esperado finito **para quaisquer** μ_0, c_0, α e β , de modo que podemos encontrar q resolvendo o seguinte sistema (pois as distribuições em Ω' ainda pertencem à família exponencial):

$$\mathbb{E}_{\mathcal{N}}[H_q] = \mathbb{E}_{\mathcal{NG}}[H_q], \quad (15)$$

Resolvendo o sistema acima obtemos os seguintes parâmetros para a posteriori Normal:

$$\begin{aligned} f_1^* &= \mu_0^* \\ f_2^* &= \psi(\alpha^*) - \ln(\beta^*) \\ q_2^* &= \psi'(\alpha^*) \\ \rho^* &= 0, \end{aligned} \quad (16)$$

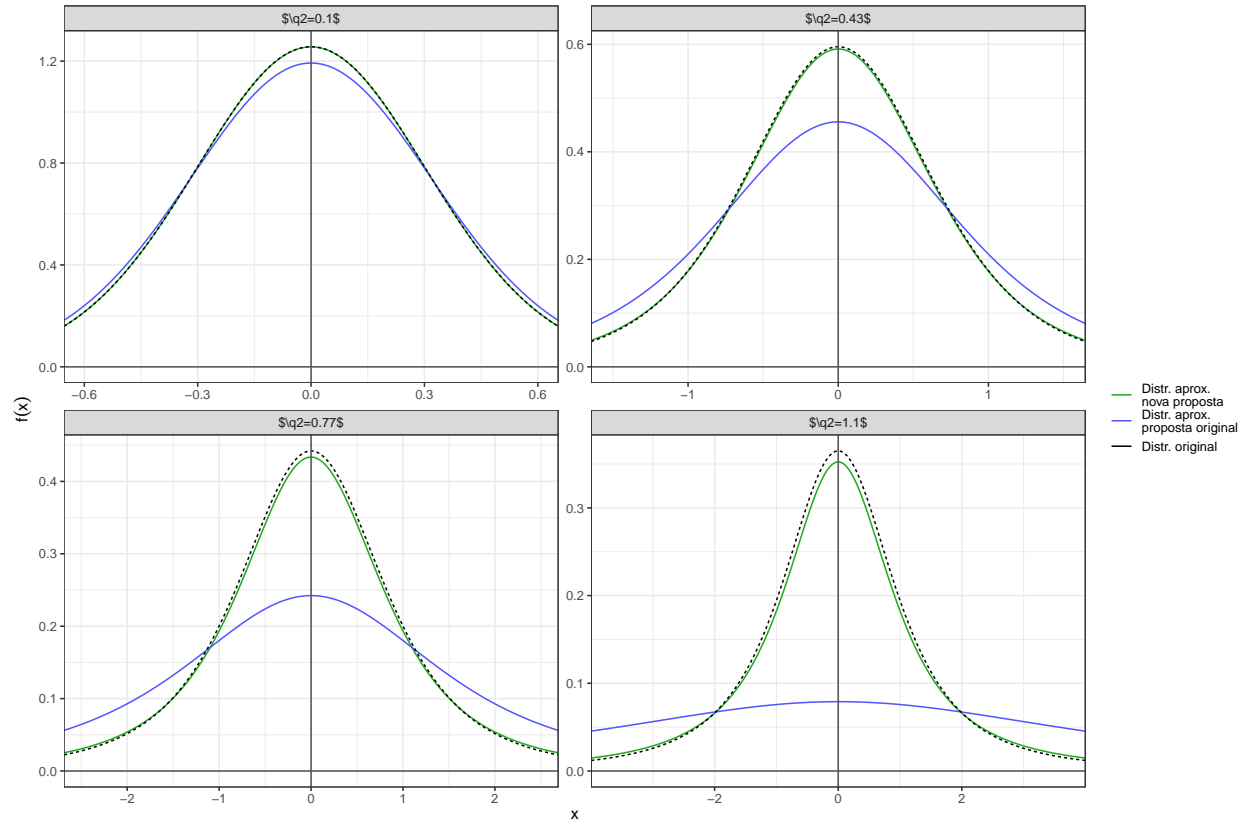
ademais, como $q \in \Omega'$, temos que $q_1 = c$.

Para completar esta proposta, resta apenas escolher o valor c . Acredito que a escolha ideal (e mais natural) seria $c = \frac{\beta}{c_0\alpha}$, pois:

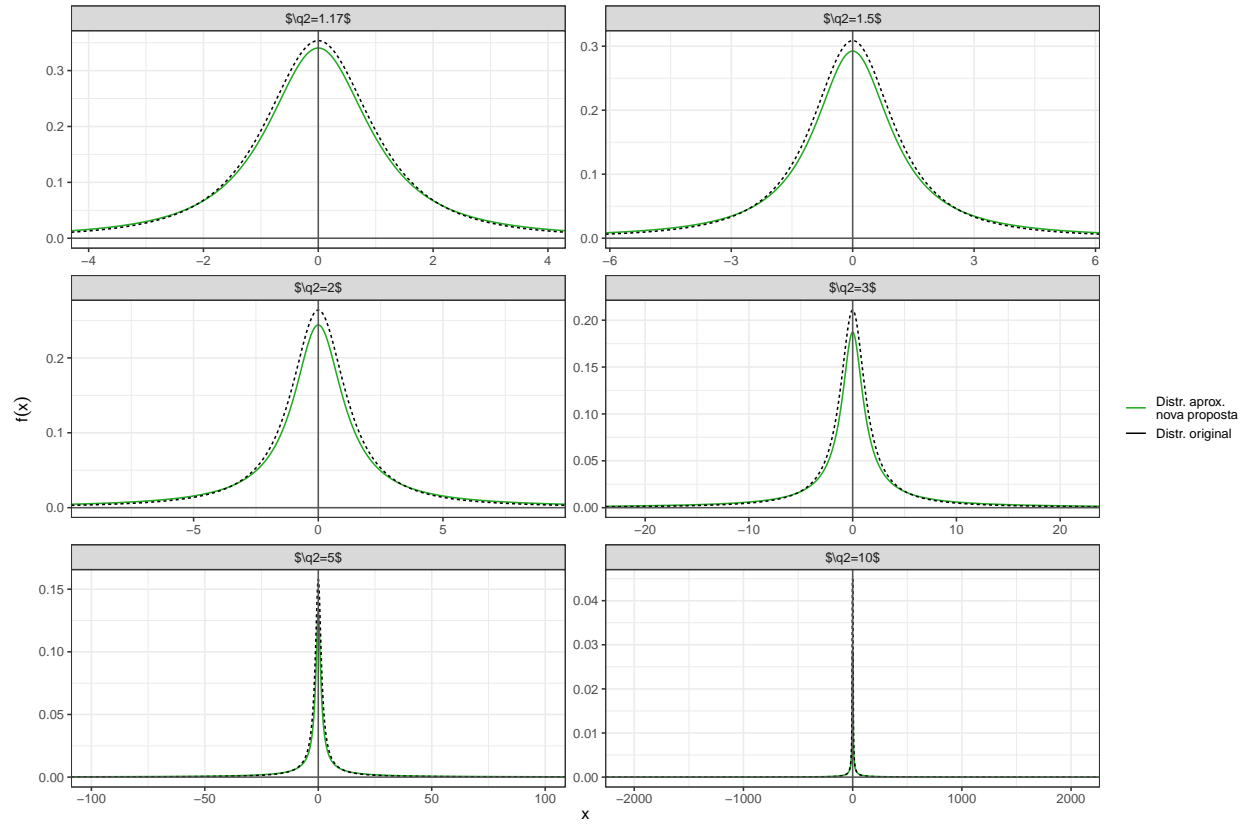
- $\frac{\beta}{c_0\alpha} \approx \frac{\beta}{c_0(\alpha-1)}$ para α grande, isto é, o mínimo em Ω' se aproxima do mínimo global da divergência KL quando esse mínimo existe, assim, não há perdas significativas na qualidade da aproximação quando α é grande;
- $\frac{\beta}{c_0\alpha}$ é bem comportado para qualquer possível valor de α . De fato, na proposta original, mesmo quando $\alpha > 1$, se α fosse próximo de 1 o ajuste ficava numericamente instável, prejudicando a qualidade da aproximação mesmo em valores válidos de α . Naturalmente, nesta nova proposta, ainda temos problemas se $\alpha \approx 0$, contudo, ao atualizarmos a nossa priori e obtermos nossa posteriori, é garantido que $\alpha \geq 0.5$, logo é garantido que a aproximação será sempre bem comportada numericamente;
- $\frac{\beta}{c_0\alpha}$ garante que não há distorção na compatibilização das distribuições, isto é, se começamos com $\lambda_1, \lambda_2 \sim \mathcal{N}_2(f_1, q_1, f_2, q_2, \rho)$, obtemos a priori conjugada \mathcal{NG} e retornamos para a priori Normal sem atualizar os parâmetros da \mathcal{NG} , é garantido que a variância na volta será idêntica a da ida;
- $\frac{\beta}{c_0\alpha}$ é o parâmetro de escala da distribuição marginal de μ , de fato, se $\mu, \phi \sim \mathcal{NG}(\mu_0, c_0, \alpha, \beta)$, então $\mu \sim t\left(2\alpha, \mu_0, \frac{\beta}{c_0\alpha}\right)$, logo, $\frac{\mu - \mu_0}{\sqrt{\frac{\beta}{c_0\alpha}}} \sim t(2\alpha)$. Assim, em essência, a minha proposta equivale a aproximar a distribuição de $\frac{\mu - \mu_0}{\sqrt{\frac{\beta}{c_0\alpha}}}$ por uma Normal Padrão.

Por último, para completar o argumento, apresentarei a qualidade da aproximação usando a nova proposta em testes análogos aos realizados com a proposta original.

Comparação da distribuição marginal de μ original e após a compatibilização:

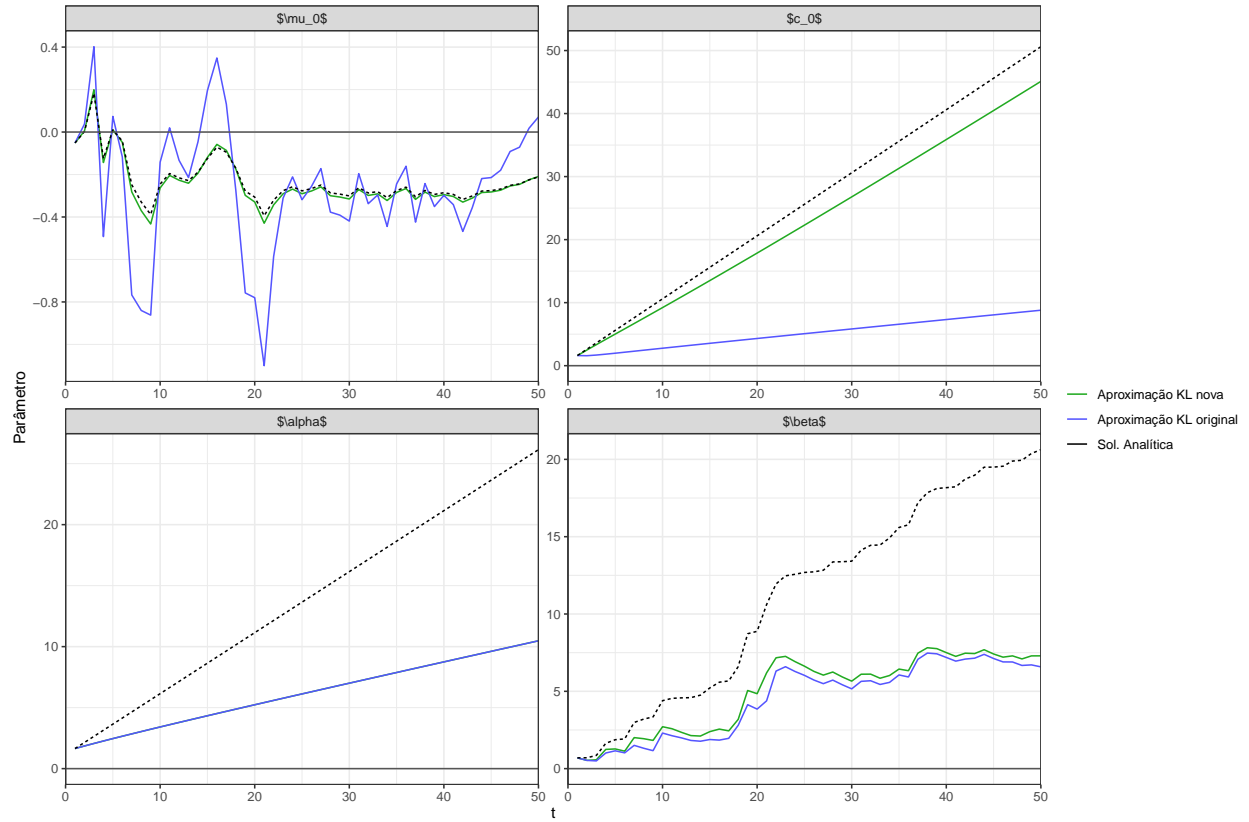


Veja que há uma melhoria drástica na qualidade da aproximação com a nova proposta, ademais, com a proposta antiga, não é possível fazer a aproximação no casos em que $q_2 \geq \frac{7}{6}$, em contrapartida, na nova proposta, podemos fazer a aproximação para qualquer valor de q_2 :



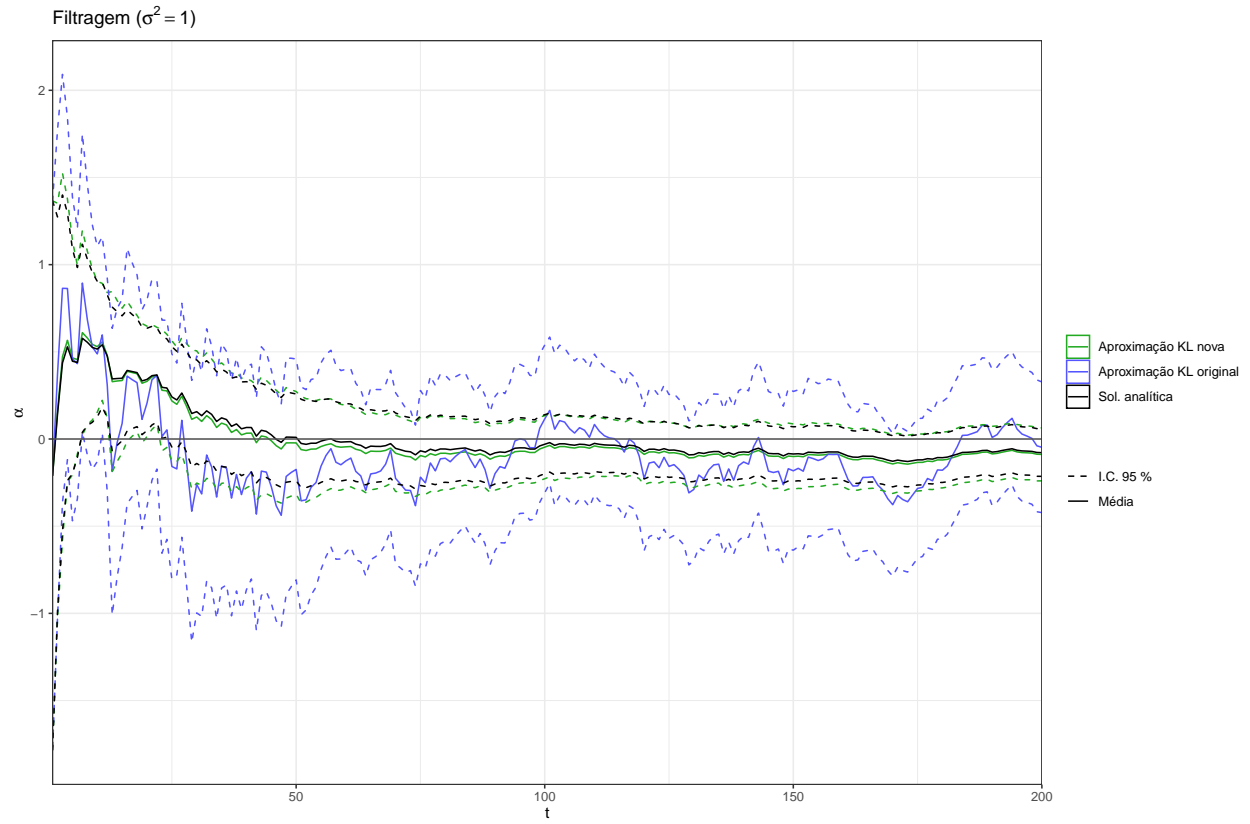
Podemos observar que a nova proposta se mostra robusta mesmo para valores bem elevados de q_2 .

Adiante comparamos a estimação dos parâmetros usando a abordagem KL com a estimação analítica para uma amostra i.i.d. de uma Normal Padrão:



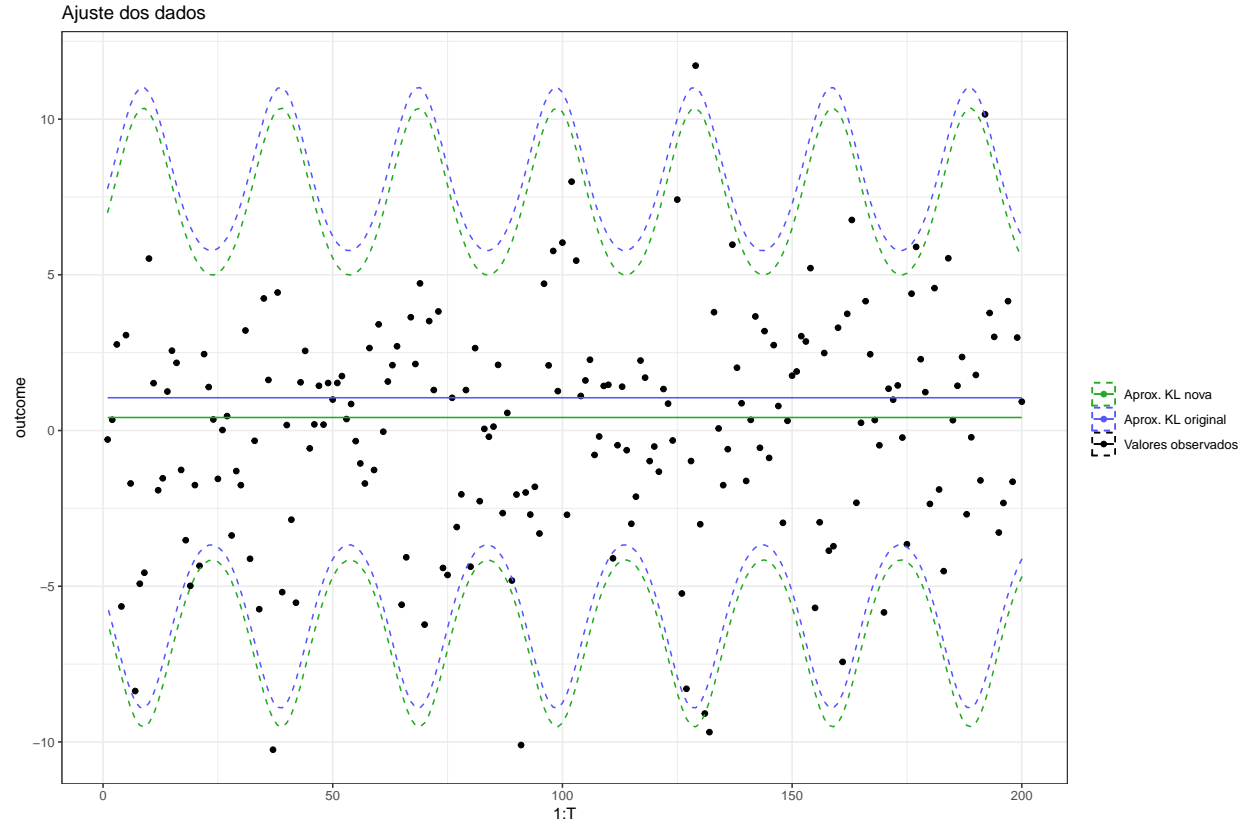
Observe que não há melhoria significativa na estimação do α e do β com a nova proposta, contudo, há um ganho significativo na estimação de μ_0 e c_0 . Vale destacar que a qualidade da aproximação da distribuição da variância será a mesma da observada no caso Gamma com parâmetro de forma igual a 0.5, assim, não há muito o que ser feito sobre os parâmetros α e β .

Olhando mais de perto para o parâmetro μ , temos a seguinte estimativa intervalar e pontual para os três ajustes para uma nova amostra de tamanho maior que a anterior:

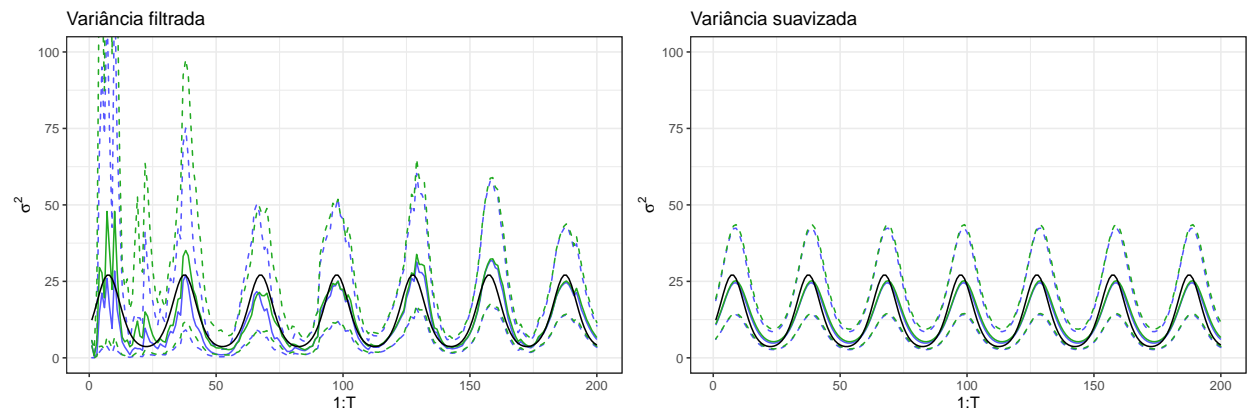


Novamente, temos que a qualidade do ajuste da média é muito superior na nova abordagem.

Em seguida vamos avaliar alguns casos onde não há solução analítica (os mesmo casos avaliados anteriormente):



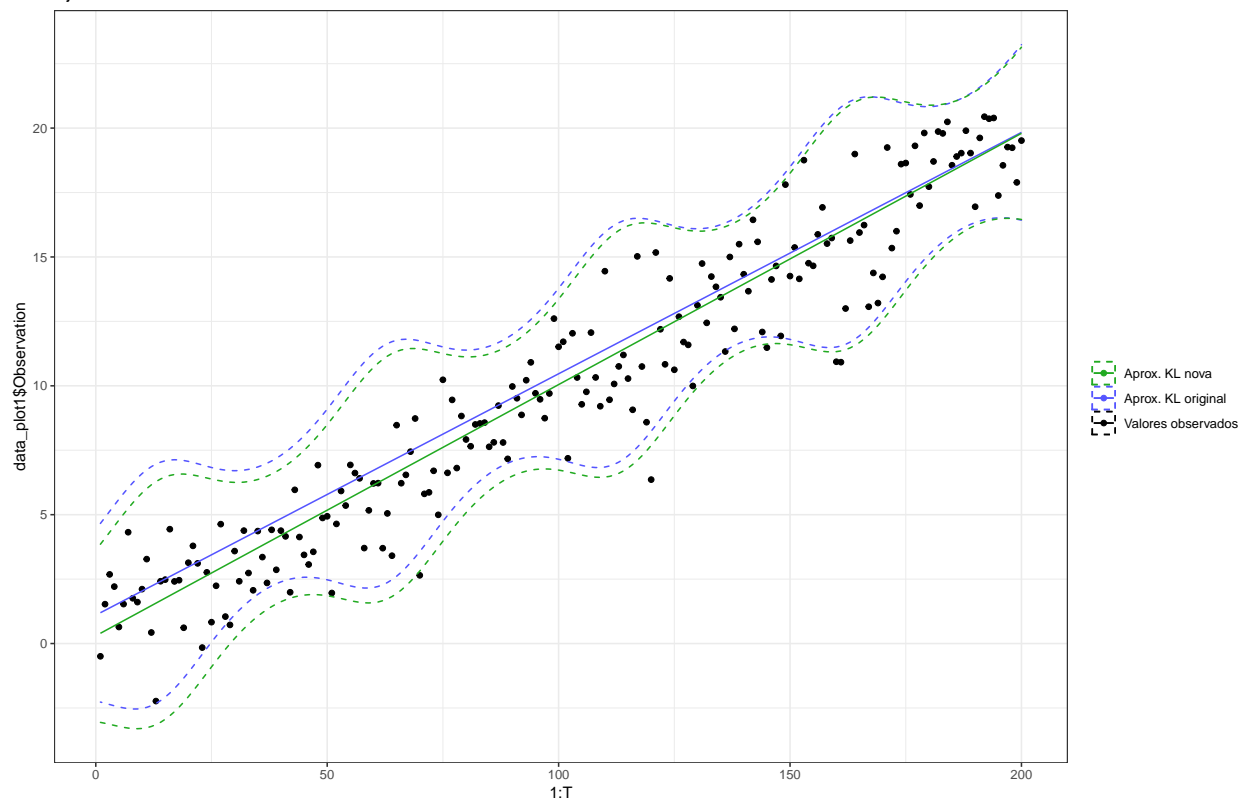
Veja que os dois modelos capturam bem o comportamento dos dados, sendo que a principal diferença entre a proposta original e a nova proposta é a estimativa da média que, na nova proposta, está mais próxima do valor correto (0). Vale destacar que a diferença entre os i.c. é essencialmente uma questão de locação, pois, como veremos adiante, as duas abordagens tem uma estimativas bem próxima para a variância:



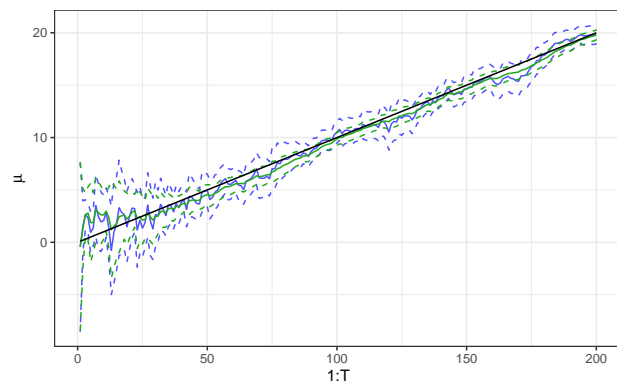
As curvas em preto representam o valor real do parâmetro em cada tempo, as curvas em azul representam o ajuste com a abordagem KL e as curvas em verde representam o ajuste com a nova proposta, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Adiante mostraremos o ajuste para uma situação parecida, mas agora com um modelo de crescimento linear para a média:

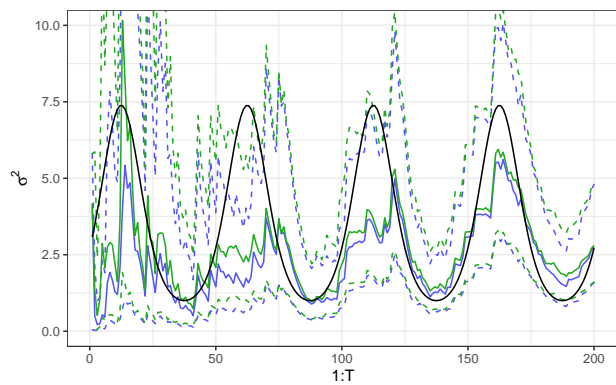
Ajuste dos dados



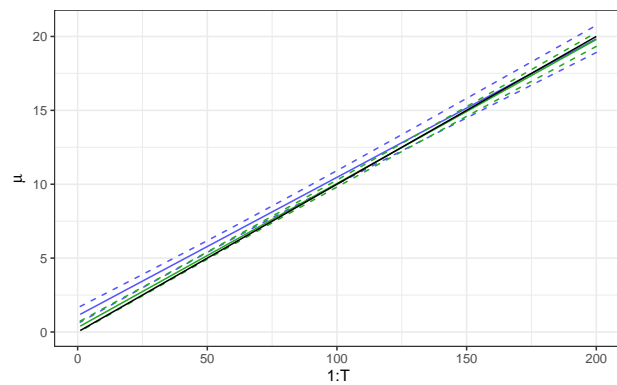
Média filtrada



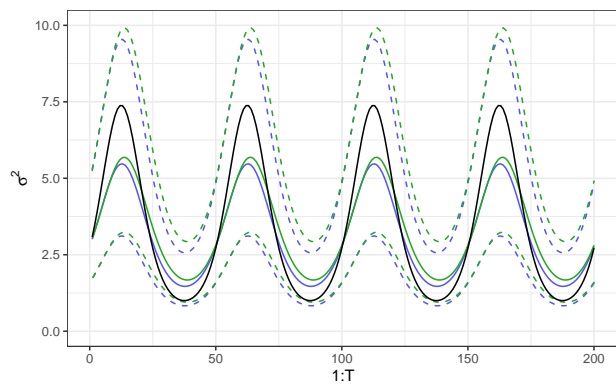
Variância filtrada



Média suavizada



Variância filtrada



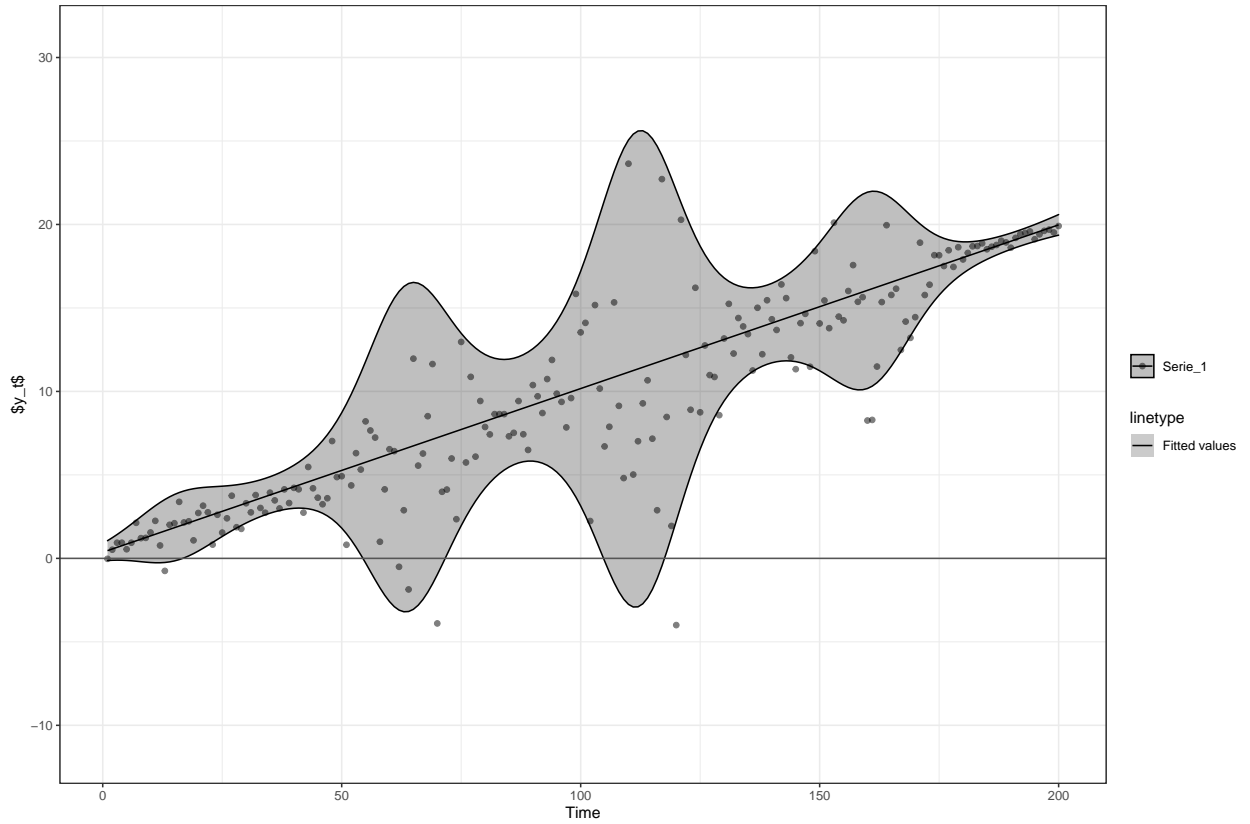
As curvas em preto representam o valor real do parâmetro em cada tempo, as curvas em azul representam o ajuste com a abordagem KL original e as curvas em verde representam o ajuste com a nova proposta, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

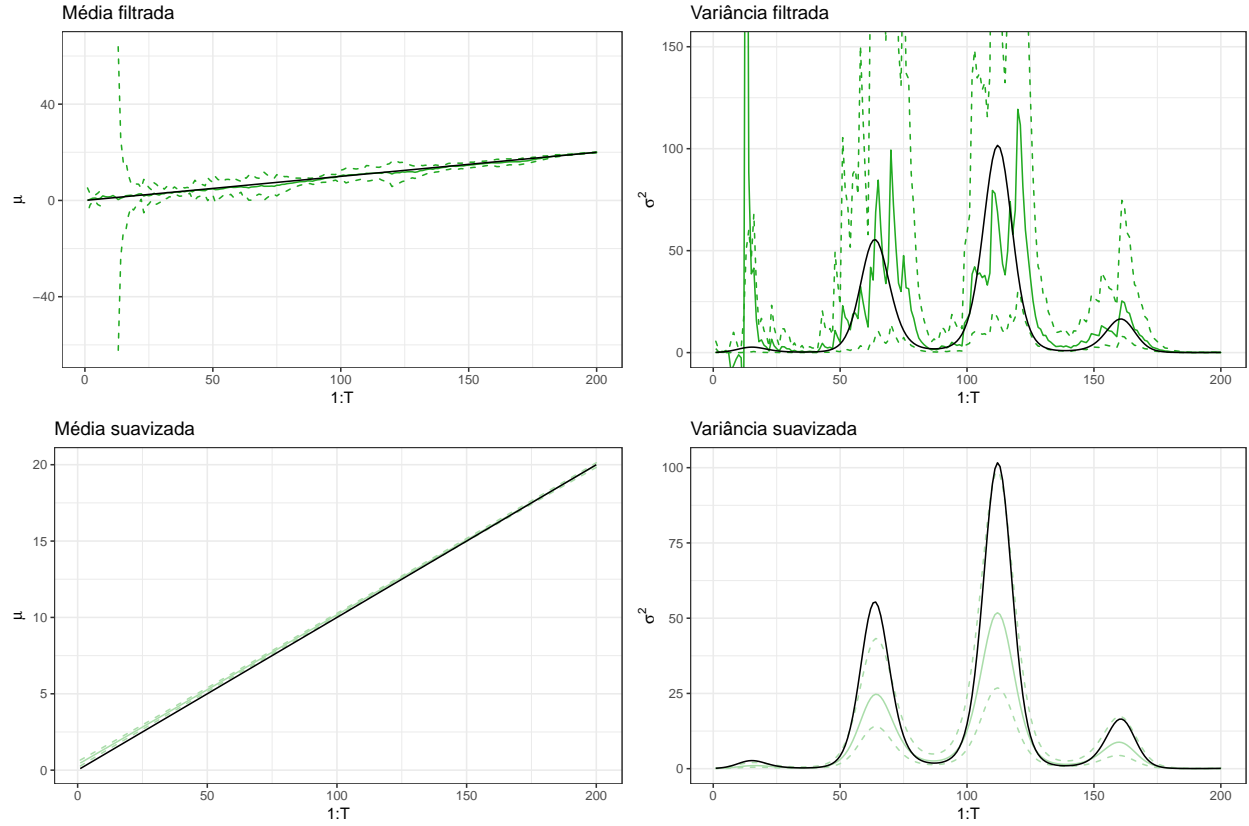
Novamente, observamos uma melhoria significativa na estimação de μ com a nova abordagem, sendo que as estimativas de ϕ tem praticamente a mesma qualidade.

Finalmente, o último exemplo a ser mostrado é um caso onde a proposta original não pode ser utilizada, i.e., um caso onde o comportamento da variância exige um modelo com estrutura mais complexa, fazendo com que o valor inicial de q_2 seja grande e a proposta original se torne inválida. Ajustaremos o modelo para dados gerado de uma Normal com média μ_i e variância S_i tais que:

$$\mu_i = \frac{i}{10}$$

$$S_i = \exp \left\{ 2 \sin \left(\frac{i\pi}{25} \right) - 20 \left(\frac{i}{20} \right) \left(\frac{i}{200} - 1 \right) - 1 \right\}$$





As curvas em preto representam o valor real do parâmetro em cada tempo e as curvas em verde representam o ajuste com a nova proposta, ademais, as curvas pontilhadas representam o i.c. de 95% e as sólidas representam as estimativas pontuais a cada tempo.

Concluimos então que a nova proposta não só produz uma aproximação superior para a média, mas também permite que a modelagem de ϕ seja feita sem restrições quanto a complexidade.