

Métricas e fatores de variabilidade na avaliação de pandemias

Paulo Hubert¹

¹ Bacharel e mestre em Estatística pelo IME-USP, Doutor em Matemática Aplicada pelo IME-USP, Professor do Departamento de Tecnologia e Ciência de Dados da EAESP/FGV

1 Introdução

Este documento pretende abrir uma discussão sobre o uso de dados e modelos para acompanhamento e comparação da disseminação do vírus SARS-CoV-1 entre diferentes regiões, países e momentos do tempo.

Entende-se que o acompanhamento da pandemia pode se dar com dois objetivos principais: estudar a dinâmica do processo de contágio (número de infectados ao longo do tempo) e estudar as consequências do processo (número de mortos, impacto na saúde pública, etc). É importante explicitar o objetivo da análise para entender quais as melhores métricas e quais as suas limitações.

A discussão será centrada na análise do modelo de crescimento exponencial simples, e se dividirá em duas partes: métricas de avaliação, e fatores de variabilidade.

Este documento não será submetido a nenhuma revista com revisão por pares, mas será divulgado abertamente para que a comunidade possa corrigi-lo e aumentá-lo. As opiniões contidas nesta primeira versão são do autor, que não fala por nenhuma instituição.

2 Processos dinâmicos e crescimento exponencial

O processo de contágio e disseminação do vírus através da população é um processo dinâmico (ocorre no tempo) que, por princípio, obedece a uma lei de crescimento exponencial.

Isto ocorre por que uma pessoa infectada pode transmitir o vírus a diversas outras pessoas, que por sua vez transmitirão o vírus a outras tantas pessoas. Se, em média, um infectado transmitir o vírus a 2 pessoas por dia, e essas 2 a outras 2 pessoas por dia, e assim por diante, em 10 dias serão 1.024 infectados; em 20 dias, serão 1.048.576; em 30 dias, mais de 1 bilhão.

O tamanho total da epidemia é limitado superiormente pelo tamanho da população (não pode haver mais infectados do que pessoas). A existência deste limite superior garante que eventualmente o processo chegará ao fim; no pior caso, isto ocorrerá quando todos houverem se contaminado e adquirido a imunidade (se existir) ou se contaminado e morrido.

A intensidade da evolução do processo depende da taxa de contágio da doença, que por sua vez pode ser decomposta em dois fatores:

- Transmissibilidade: considerando que um infectado entre em contato com uma pessoa saudável, qual a probabilidade de ocorrer o contágio?
- Exposição: no tempo entre o contágio e a cura ou morte, quantas pessoas entrarão em contato (direto ou indireto) com o infectado?

A transmissibilidade depende principalmente de fatores biológicos: a agressividade do vírus, seu tempo de sobrevivência em diferentes condições, seu tempo de incubação, a intensidade dos sintomas, entre outros. Mas pode também depender de fatores culturais; por exemplo, da distância usual que as pessoas de determinada cultura guardam umas das outras ao conversar, hábitos de higiene, etc.

A exposição, por outro lado, depende principalmente de fatores culturais: a conectividade do território e a intensidade da vida social são dois fatores importantes. Mas também depende de fatores biológicos, como por exemplo o tempo até o aparecimento dos sintomas (quanto maior este tempo, mais pessoas entrarão em contato com o infectado).

3 As consequências

A consequência mais severa da contaminação é a morte do doente; conhecer a probabilidade de que um indivíduo doente venha a falecer (a taxa de mortalidade ou letalidade) é portanto essencial para estimar o dano causado pelo processo ao longo de seu curso.

Há, porém, uma série de outras consequências do processo de contágio; no contexto da saúde pública, há o aumento no nível de ocupação das estruturas de saúde (postos, hospitais) e o aumento rápido da demanda por serviços, profissionais e equipamentos de saúde.

Diversos outros efeitos deste processo podem ser observados, desde efeitos econômicos pela redução na velocidade de produção e consumo, até efeitos sociais pelo estresse causado pela própria pandemia ou pelas medidas de contenção.

4 Comparando processos exponenciais

4.1 Métricas

4.1.1 Métricas para avaliar a dinâmica de contágio

Para entender melhor quais métricas são ou não adequadas, vamos analisar as equações de um processo exponencial simples.

O número de infectados I_{t+1} num dia qualquer $t + 1$ depende do número de infectados no dia anterior, I_t ; no caso do processo exponencial, esta dependência é multiplicativa, ou seja

$$I_{t+1} = \alpha I_t \tag{1}$$

O valor de α governa o comportamento qualitativo do processo; se $\alpha < 1$, o número de infectados diminui com o tempo e chegará finalmente a 0 (considerando que $I_t \in \mathbb{N}$). Caso

$\alpha > 1$, o número de infectados crescerá sem parar (ainda não estamos considerando o limite do tamanho da população).

O processo descrito na equação acima é dito exponencial pela seguinte razão: se soubermos o número de infectados num dia qualquer I_0 (por exemplo $I_0 = 1$ na data inicial), podemos calcular o número de infectados no dia t da seguinte maneira:

$$I_t = I_0 \cdot \alpha^t \quad (2)$$

Ou seja, o tempo transcorrido t aparece como expoente na fórmula que dá o número de infectados ao longo do tempo.

Nesta equação, o coeficiente principal é a base α da exponenciação. Como vimos acima, seu valor determina qualitativamente o comportamento do processo (havendo uma mudança importante entre $\alpha < 1$ e $\alpha > 1$); mas também determina a dinâmica quantitativamente, pois está relacionado à velocidade de crescimento em cada instante de tempo t (pois em processos exponenciais a velocidade de crescimento não é constante).

Sendo assim, o valor de α em diferentes países seria uma boa métrica de comparação de acordo com o modelo 2.

Uma estatística comum para avaliar α indiretamente é o *doubling time*, ou tempo para dobrar. Esta estatística é dada em unidades de tempo e representa o tempo estimado para que o número de infectados dobre. Usando a equação 2, queremos encontrar k tal que $I_{t+k} = 2I_t$; assim

$$\begin{aligned} I_{t+k} &= I_0 \alpha^{t+k} = 2I_t = 2I_0 \alpha^t \\ \iff \alpha^{t+k} &= 2\alpha^t \iff \alpha^k = 2 \\ \iff k &= \frac{\log(2)}{\log(\alpha)} = \log_\alpha 2 \end{aligned}$$

Há portanto uma relação direta entre o tempo para dobrar e o valor de α ; o tempo para dobrar, por ser mais interpretável, é uma alternativa de estatística para a comparação de processos exponenciais.

Outra possibilidade é a comparação de estimativas diretas de α ; estas estimativas podem ser obtidas de diferentes maneiras, e envolvem de modo geral o ajuste do modelo 2 aos dados do número de infectados no tempo.

Por outro lado, o modelo 2 não descreve bem a evolução de uma pandemia concreta; segundo este modelo, se $\alpha > 1$ a pandemia crescerá indefinidamente, o que não é possível pelo tamanho finito da população. O mais correto, portanto, é aumentar este modelo e torná-lo mais preciso.

Em epidemiologia isto costuma ser feito através de modelos compartimentados do tipo *SIR* ou *SEIR*. Estes modelos dividem a população total em subgrupos: *Suscetíveis*, *Expostos*, *Infectados* e *Recuperados* (os mortos são contabilizados como recuperados, no sentido de não poderem mais se tornar vetores da doença), e consideram que o tamanho N da população total é fixo e dado por $N = S + E + I + R$ ¹. Estes modelos incluem diversos outros

¹Há outros modelos compartimentados que fazem divisões mais detalhadas, mas a discussão de toda esta classe de modelos vai além do escopo deste texto.

parâmetros além de α e I_0 para descrever a dinâmica temporal entre os grupos (suscetíveis se tornar expostos com uma certa velocidade, expostos tornam-se infectados com uma certa probabilidade, e infectados recuperam-se após um certo tempo), e os valores estimados destes parâmetros podem ser usados como métricas de avaliação e acompanhamento.

Algumas métricas são inaceitáveis. Por exemplo, o número de novos casos ou de mortes em um período de tempo não é uma informação suficiente, pois dois países com números iguais de novos casos podem estar vivendo realidades muito diferentes. 100 casos a mais num país com 1000 casos atualmente confirmados representa um contexto bem diferente de 100 casos a mais num país com 10 casos atualmente confirmados.

A métrica mais simples aceitável é o fator de crescimento I_{t+1}/I_t , que representa uma estimativa do valor de α (assumindo sem perda de generalidade que $I_0 = 1$) no modelo 2. Este valor, conforme varia no tempo, agrega as consequências de todas as outras dinâmicas presentes no modelo mais completo. Por exemplo, no momento em que for detectada uma diminuição importante nos valores diários estimados desta taxa, poderemos acreditar que o processo se aproxima do fim, e / ou que as medidas preventivas estão surtindo efeito.

Métricas mais complexas exigem um esforço computacional que nem sempre estará à disposição dos interessados em acompanhar a situação. Quando possível, porém, devem ser preferidos, desde que o modelo utilizado na sua estimação seja claramente exposto.

É importante ressaltar, porém, que toda métrica calculada a partir do número de casos está necessariamente sujeita à incerteza, pois o processo de rastreamento, testagem e contabilização de casos está sujeito a falhas e imprecisões.

4.1.2 Métricas para avaliar consequências da pandemia

As consequências da pandemia estão relacionadas principalmente ao número de mortes e à intensidade dos sintomas (que indicarão a necessidade de UTI, respiradores e outros equipamentos).

Para projetar o número de mortes, a estatística mais comum é a taxa de letalidade, que pode ser definida de duas formas: a proporção de mortos no *total da população*, e a proporção de mortos no *total de infectados*.

O processo de contágio e disseminação de um vírus é um processo essencialmente local, pois a transmissão ocorre pelo contato direto ou indireto. Sendo assim, a taxa de letalidade pelo total de infectados pode ser mais informativa, já que mede mais diretamente o poder destrutivo do vírus. Por outro lado, esta métrica é mais incerta, pois o cálculo do número de infectados está sujeito a incertezas por si só. O mais importante na avaliação desta taxa é garantir a consistência (todas as taxas comparadas devem ser calculadas no mesmo critério), e a exposição clara para o leitor de qual taxa está sendo usada, e porquê.

Quanto à intensidade dos sintomas, não parece razoável definir uma única métrica ou utilizar um único número para sua avaliação. Muitas métricas podem ser propostas: tempo até aparecimento do primeiro sintoma, probabilidade do paciente precisar de cuidados hospitalares, probabilidade de necessidade de UTI, tempo esperado de internação, entre outros. Todas essas métricas serão potencialmente influenciadas por características de cada paciente, como idade, estado de saúde, etc. Sendo assim, é preciso cuidado na comparação destas métricas.

A análise das consequências econômicas ou sociais da pandemia ou das medidas de prevenção está além do escopo desta primeira versão do relatório.

4.2 Fatores de variabilidade

No momento atual, a maioria dos países do mundo já apresenta casos positivos do SARS-CoV-1. O vírus que afeta cada um dos países é essencialmente o mesmo; há motivos para crer, portanto, que os fatores biológicos envolvidos na evolução e consequências da pandemia serão constantes ao redor do planeta.

Por outro lado, a diversidade cultural e de infraestrutura econômica garante que cada país terá um processo diferente dos demais, em maior ou menor grau.

Para comparar a evolução da pandemia e suas consequências entre diferentes países é preciso levar em conta suas diferenças; mas é preciso também saber onde essas diferenças podem ser ignoradas, onde podem ser incorporadas na análise para tornar a comparação mais precisa, e onde elas impedem completamente uma comparação honesta.

Listo abaixo alguns fatores, sem pretensão de esgotá-los todos.

4.2.1 Data de início

A data de início do processo de contágio pode ser definida de diversas maneiras: pela data em que o primeiro infectado entrou no território, pela data em que apresentou os primeiros sintomas, pela data em que a presença da doença foi confirmada por teste clínico.

Seja qual for a definição escolhida, para avaliar a situação de países diferentes é preciso considerar o estágio do processo em que cada país se encontra. Em outras palavras, comparar qualquer métrica entre dois países na mesma data não é correto.

Esta diferença, porém, pode ser eliminada da análise caso se considere a evolução do processo tomando como $t = 0$ a data de início. Em outras palavras, o tempo deve ser medido não pela data do calendário, mas pelo tempo transcorrido desde o início do processo.

Há diversas imprecisões envolvidas na determinação desta data inicial, o que torna a comparação entre países uma tarefa mais difícil e sujeita ao erro e à incerteza. Como aproximação razoável, porém, cabe utilizar a normalização do tempo conforme descrito acima.

4.2.2 Tamanho da população

O tamanho total da população serve como limitante superior do número total de infectados. Sendo assim, e considerando que o processo de contágio é local (ou seja, o vírus é transmitido apenas pelo contato direto ou indireto), a diferença no tamanho total da população é pouco relevante, principalmente no início da pandemia. Neste início são muito mais relevantes a conectividade entre os cidadãos e suas relações sociais, que tem efeito local.

Há alguns artigos propondo a utilização do número de casos por milhão de habitantes ou métricas semelhantes. Na minha opinião esta proposta não está correta, a menos que se queira medir o tamanho do impacto da pandemia na estrutura do país. Para comparar diferentes processos ao longo do tempo, a métrica mais importante é a taxa de variação diária, que não depende do tamanho da população.

4.2.3 Pirâmide etária

Uma vez que os efeitos da infecção pelo SARS-CoV-1 dependem criticamente do estado de saúde e da idade do paciente, a estrutura etária da população é um fator importante. É

preciso considerar, porém, como e de que modo essa estrutura influencia a análise.

Na análise do processo de contágio e sua velocidade, a estrutura etária é importante à medida que a idade do paciente interfira *a)* no tempo até o aparecimento dos sintomas e *b)* na intensidade da convivência social do infectado. Portanto, ao se comparar diferentes países quanto à evolução do processo de contágio, é importante considerar se as diferenças nas pirâmides etárias podem ter influência muito grande nestes dois fatores.

Na análise das consequências do processo, a estrutura etária é importante à medida que vai influenciar *a)* o número de mortos, e *b)* a carga sobre o sistema de saúde. Sendo assim, não se pode comparar diretamente países com estruturas etárias muito diferentes no que diz respeito ao impacto da pandemia na saúde pública.

É possível mitigar os efeitos da diferença nas pirâmides etárias utilizando-se indicadores adequados; por exemplo, pode-se observar o número de mortos ou infectados entre a população de uma determinada faixa etária, ou pode-se proceder a diversos tipos de normalização. A abordagem formalmente mais correta seria incluir esta informação no modelo utilizado para descrever o processo. O modo adequado de fazer isso depende do tipo de modelo.

4.2.4 Estrutura e organização social

A estrutura e organização social de um país podem ser decisivas na determinação da evolução e consequências de um processo pandêmico. A disponibilidade de recursos e equipamentos de saúde, a capacidade de mobilizar novos recursos rapidamente, a capacidade de comunicação do governo e a competência na implantação de políticas de contenção são alguns fatores que se mostram potencialmente influentes.

Comparar países com estruturas ou organizações sociais muito diferentes (Alemanha e Brasil, por exemplo, ou Brasil e China) é um procedimento complexo. É preciso considerar o peso relativo das semelhanças (o fato de ser o mesmo vírus, e outras que pode haver) e diferenças entre os países para avaliar criticamente a qualidade de uma comparação deste tipo.

Novamente, é possível em tese mitigar os efeitos dessa diferença usando as ferramentas adequadas de análise. Estas ferramentas, porém, exigirão maior tempo de análise e processamento, e em situações normais os resultados assim obtidos devem passar pelo crivo da análise por pares antes de ir à público.

4.3 Conclusão

O objetivo deste texto era abrir a discussão sobre métricas adequadas ao acompanhamento da evolução da pandemia do SARS-CoV-1, e sobre os fatores de variabilidade que permitem ou impedem a comparação entre países. Pretende-se que esta discussão esteja acessível a um público o mais vasto possível. Infelizmente, não é possível eliminar completamente os detalhes técnicos e matemáticos sem sacrificar por demais a exatidão e a clareza lógica.

De forma resumida, esta primeira versão propõe que a métrica mais adequada para acompanhamento da dinâmica é a taxa de crescimento diário do número de casos confirmados, I_{t+1}/I_t . A taxa de crescimento diário do número de mortes é também uma métrica adequada, que pode estar sujeita a menos fontes de erro do que a determinação do número de

contaminados.

Para avaliação das consequências da pandemia, é essencial considerar a diferença de infraestrutura dos países; a diferença entre as datas de início podem ser eliminadas por um processo de normalização, e outras diferenças como na pirâmide etária podem ter seu efeito mitigado na análise. O tamanho total das populações não impede as comparações, principalmente no início da pandemia e se consideramos que o processo de contágio é local.