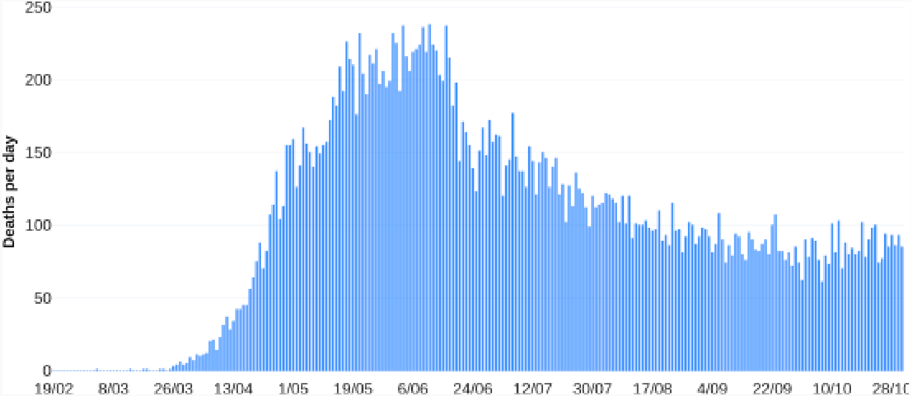
Estimación máximo verosímil para un sistema SEIR estocástico

con una aplicación COVID-19

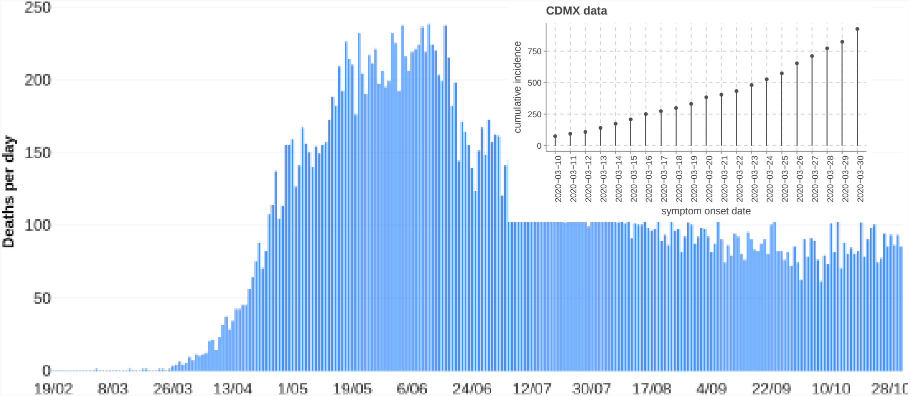
October 25, 2023, 56 Congreso SMM

CONAHCYT-UNISON-UNAM: FDV, FBL, SDIV

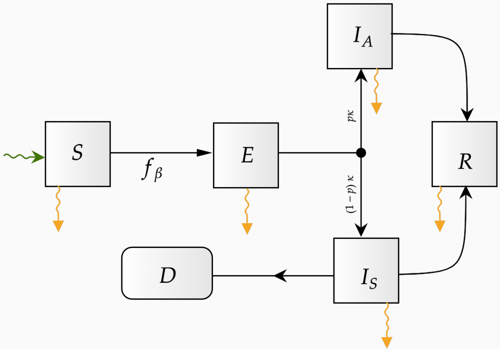
CDMX Covid-19 data



CDMX Covid-19 data



First attempt: MCMC with a deterministic SEIRS structure



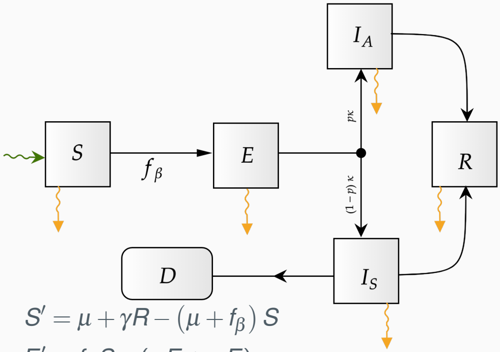
$$f_\beta := \beta_s I_s + \beta_a I_a$$

$$\frac{S + E + I + R + D}{N} = 1$$

$$\beta_\bullet = \frac{\widehat{\beta}_\bullet}{N - D}$$

μ

First attempt: MCMC with a deterministic SEIRS structure



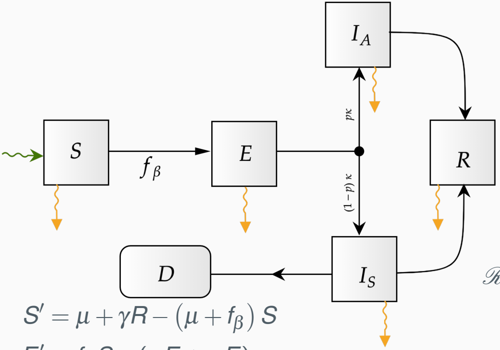
$$f_\beta := \beta_s I_s + \beta_a I_a$$

$$\frac{S+E+I+R+D}{N} = 1$$

$$\beta_\bullet = \frac{\widehat{\beta}_\bullet}{N-D}$$

$$S' = \mu + \gamma R - (\mu + f_\beta)\, S$$
$$E' = f_\beta S - (\kappa E + \mu E)$$
$$I_a' = p\kappa E - (\alpha_a + \mu)\, I_a$$
$$I_s' = (1-p)\kappa E - (\alpha_s + \mu) I_s$$
$$R' = \alpha_a I_a + \alpha_s (1-\theta) I_s - (\mu + \gamma) R$$
$$D' = \theta \alpha_s I_s.$$

First attempt: MCMC with a deterministic SEIRS structure



$$f_\beta := \beta_s I_s + \beta_a I_a$$

$$\frac{S+E+I+R+D}{N} = 1$$

$$\beta_\bullet = \frac{\widehat{\beta}_\bullet}{N-D}$$

μ

$$\mathscr{R}_0^D := \frac{p\kappa\beta_s}{(\mu+\kappa)(\mu+\alpha_s)} + \frac{(1-p)\kappa\beta_a}{(\mu+\kappa)(\mu+\alpha_a)}.$$

$$
\begin{aligned}
S' &= \mu + \gamma R - (\mu + f_\beta)\, S \\
E' &= f_\beta S - (\kappa E + \mu E) \\
I_a' &= p\kappa E - (\alpha_a + \mu)\, I_a \\
I_s' &= (1-p)\kappa E - (\alpha_s + \mu) I_s \\
R' &= \alpha_a I_a + \alpha_s (1-\theta) I_s - (\mu + \gamma) R \\
D' &= \theta \alpha_s I_s.
\end{aligned}
$$

$$
\begin{aligned}
Y_t &\sim \mathrm{Poisson}(\lambda_t) \\
\lambda_t &= \int_0^t (1-p)\kappa E \\
p &\sim \mathrm{Uniform}(0.3, 0.8) \\
\kappa &\sim \mathrm{Gamma}(10, 50) \\
\beta_a, \beta_s &\sim \mathscr{N}(0.5, 0.1)
\end{aligned}
$$

Parameter values of the deterministic model

Parameter	Value	Reference
μ^{-1}	70×365	[1]
β_s	0.05821	Estimated
β_a	0.510968	Estimated
κ	0.196078^{-1}	[2]
p	0.585505	Estimated
θ	0.11	[1]
α_s	0.092507^{-1}	[1]
α_a	0.167504^{-1}	[1]
γ^{-1}	365	[1]

Table: Parameter values of the model

Manuel Adrian Acuña-Zegarra, Saúl Díaz-Infante, David Baca-Carrasco, and Daniel Olmos-Liceaga.
COVID-19 optimal vaccination policies: A modeling study on efficacy, natural and vaccine-induced immunity responses.
Mathematical Biosciences, 337:108614, 2021.

Huaiyu Tian, Yonghong Liu, Yidan Li, Chieh-Hsi Wu, Bin Chen, Moritz U. G. Kraemer, Bingying Li, Jun Cai, Bo Xu, Qiqi Yang, Ben Wang, Peng Yang, Yujun Cui, Yimeng Song, Pai Zheng, Quanyi Wang, Ottar N. Bjornstad, Ruifu Yang, Bryan T. Grenfell, Oliver G. Pybus, and Christopher Dye.
An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china.
Science, 368(6491):638–642, 2020.

Overfiting example

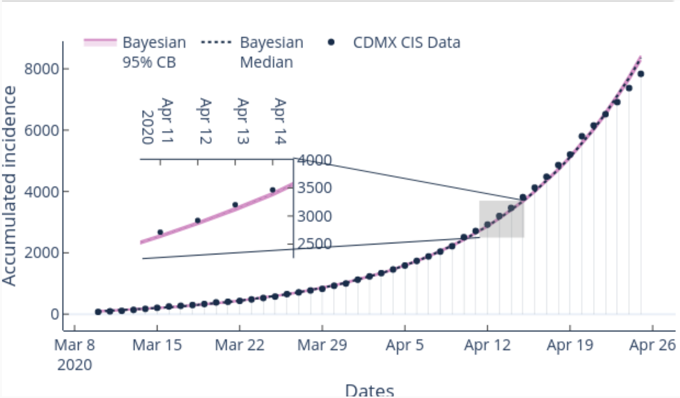


Figure: MCMC Fit of diary new cases of Mexico city during exponential growth. See `https://plotly.com/~sauld/53/` for an electronic version.

Aims, methodology, and contribution

Aim: Improve the estimation of the infection rates β_s, β_a, and ratio of asymptomatic cases p.

Hypothesis: Noise could improve the uncertainty quantification.

Aim: Improve the estimation of the infection rates β_s, β_a, and ratio of asymptomatic cases p.

Hypothesis: Noise could improve the uncertainty quantification.

Solution idea:

- Deterministic Inference
  - ✓ ODE
  - ✓ MCMC
- Stochastic Inference
  - ❑ SDE extension
  - ❑ Lamperti
  - ❑ Grisanov
  - ❑ Opt. ML

Aims, methodology, and contribution

Aim: Improve the estimation of the infection rates β_s, β_a, and ratio of asymptomatic cases p.

Hypothesis: Noise could improve the uncertainty quantification.

Solution idea:

- Deterministic Inference
  - ✓ ODE
  - ✓ MCMC
- Stochastic Inference
  - ❑ SDE extension
  - ❑ Lamperti
  - ❑ Grisanov
  - ❑ Opt. ML

Contribution

- → A stochastic SEIR.
- → Consistent estimators.
- → Simulation.
- → Real data.

Our Contribution [Baltazar-Larios et al., 2022]

Baltazar-Larios, F., Delgado-Vences, F., and Diaz-Infante, S. (2022).
Maximum likelihood estimation for a stochastic SEIR system with a COVID-19 application.
International Journal of Computer Mathematics, pages 1–23.

Related Work

Works that report estimation for SDE models for similar structures (SIS,or SIR) and with only numeric experiments of synthetic nature. Complex stochastic models similar to the one considered here that are only studied on a theoretical basis and/or with numerical simulations but without estimations methods.

- [1] focus on the calibration parameters by adaptive MCMC and extended Kalman filter methods.
- [2] report other Bayesian techniques,
- [4] apply a maximum likelihood method and
- [3] reports a parameter estimation with local generalized methods of moments.

Ndanguza D., S. Mbalawata I., and P. Nsabimana J.
Analysis of SDEs Applied to SEIR Epidemic Models by Extended Kalman Filter Method.
Applied Mathematics, 07(17):2195–2211, 2016.

LUIZ K. HOTTA.
Bayesian melding estimation of a stochastic seir model.
Mathematical Population Studies, 17(2):101–111, 2010.

Olusegun M. Otunuga.
Estimation of epidemiological parameters for COVID-19 cases using a stochastic SEIRS epidemic model with vital dynamics.
Results in Physics, 28:104664, 2021.

Andrés Rıos-Gutiérrez, Viswanathan Arunachalam, and Anuj Mubayi.

Stochastic anlysis and statistical inference for seir models of infectious diseases.

Table of contents

Stochastic perturbation

Perturbing the above deterministic base by Brownian Motion

$\mu dt \rightsquigarrow \mu dt + \sigma dW(t)$ gives our SDE SEIR-Covid-19

$$
\begin{aligned}
dS(t) &= \left[\mu - \mu S(t) - f_\beta S(t) + \gamma R(t)\right] dt + \sigma\left(1 - S(t)\right) dW(t) \\
dE(t) &= \left[f_\beta S(t) - \kappa E(t) - \mu E(t)\right] dt - \sigma E(t) dW(t) \\
dI_a(t) &= \left[p\kappa E(t) - (\alpha_a + \mu) I_a(t)\right] dt - \sigma I_a(t) dW(t) \\
dI_s(t) &= \left[(1-p)\kappa E(t) - (\alpha_s + \mu) I_s(t)\right] dt - \sigma I_s(t) dW(t) \\
dR(t) &= \left[\alpha_a I_a(t) + \alpha_s I_s(t) - (\mu + \gamma) R(t)\right] dt - \sigma R(t) dW(t), \\
& \quad t \in [0, T].
\end{aligned}
$$

Grisanov's likelihood ratio

Let $\mathbb{P}_{\beta,p}$ the law of solution to SDE. We use the following result [1].

Theorem (Likelihood ratio of Itô processes Särkkä and Solin (2019, Thm. 7.4))

Consider the Itô processes

$$\begin{aligned} dx &= f(x,t) + dB_t, \qquad x(0) = x_0, \\ dy &= g(y,t) + dB_t, \qquad y(0) = x_0. \end{aligned}$$

Then the ratio of probability laws of $\mathscr{X}_t$ and $\mathscr{Y}_t$ is given as

$$\begin{aligned} \frac{p(\mathscr{X}_t)}{p(\mathscr{Y}_t)} &= Z(t), \\ Z(t) = \exp\Big(-\frac{1}{2}\int_0^t & [f(y,\tau) - g(y,\tau)]^\top \mathbb{Q}^{-1} [f(y,\tau) - g(y,\tau)] d\tau \\ &+ \int_0^t [f(y,\tau) - g(y,\tau)]^\top \mathbb{Q}^{-1} dB_\tau \Big) \end{aligned}$$

in the sense that for an arbitrary functional $h(\cdot)$ of the path from 0 to t,

$$\mathbb{E}[h(\mathscr{X}_t)] = \mathbb{E}[Z(t)h(\mathscr{Y}_t)]$$

[1] Särkkä, Simo; Solin, Arno, Applied stochastic differential equations. Institute of Mathematical Statistics Textbooks, 10. Cambridge University Press, Cambridge, 2019. ix+316 pp. ISBN: 978-1-316-64946-6

The Lamperti transform

Suppose we have the SDE

$$dX_t = a(t, X_t)dt + b(X_t)dW_t,$$

where the diffusion coefficient depends only on the state variable. Such SDE can transform into one with unitary diffusion by applying the *Lamperti* transform

$$Y_t := F(X_t) = \int_z^{X_t} \frac{1}{b(u)} du.$$

Here z is an arbitrary value and Y_t solves

$$dY_t = \left(\frac{a(t, X_t)}{b(X_t)} - \frac{1}{2} b_x(X_t) \right) dt + dW_t$$

The results follows form the Itô formula.

Using Itô and Lamperti transformations

$$
\begin{aligned}
dS(t) &= \left[\mu - \mu S(t) - f_\beta S(t) + \gamma R(t)\right] dt + \sigma\left(1 - S(t)\right) dW(t)\\
dE(t) &= \left[f_\beta S(t) - \kappa E(t) - \mu E(t)\right] dt - \sigma E(t) dW(t)\\
dI_a(t) &= \left[p\kappa E(t) - (\alpha_a + \mu) I_a(t)\right] dt - \sigma I_a(t) dW(t)\\
dI_s(t) &= \left[(1-p)\kappa E(t) - (\alpha_s + \mu) I_s(t)\right] dt - \sigma I_s(t) dW(t)\\
dR(t) &= \left[\alpha_a I_a(t) + \alpha_s I_s(t) - (\mu + \gamma) R(t)\right] dt - \sigma R(t) dW(t),\\
&\quad t \in [0, T].
\end{aligned}
$$

$$
-\tfrac{1}{\sigma} d\mathbf{X}_{\beta,p}(t) = F\left(\mathbf{X}_{\beta,p}(t)\right) dt + d\mathbf{W}(t),
$$

$$
\mathbf{X}_{\beta,p}(t) := \begin{pmatrix} \log\left(1 - S(t)\right) \\ \log\left(E(t)\right) \\ \log\left(I_a(t)\right) \\ \log\left(I_s(t)\right) \\ \log\left(R(t)\right) \end{pmatrix}, \quad
F(\mathbf{X}_{\beta,p}(t)) := \begin{pmatrix}
\frac{\mu}{\sigma} - \frac{f_\beta S(t)}{\sigma(1 - S(t))} + \frac{\gamma R(t)}{\sigma(1 - S(t))} + \frac{1}{2}\sigma \\
-\frac{f_\beta S(t)}{\sigma E(t)} + \frac{\kappa}{\sigma} + \frac{\mu}{\sigma} + \frac{1}{2}\sigma \\
-\frac{\kappa p E(t)}{\sigma I_a(t)} + \frac{(\alpha_a + \mu)}{\sigma} + \frac{1}{2}\sigma \\
-\frac{\kappa(1-p)E(t)}{\sigma I_s(t)} + \frac{(\alpha_s + \mu)}{\sigma} + \frac{1}{2}\sigma \\
-\frac{\alpha_a I_a(t) + \alpha_s I_s(t)}{\sigma R(t)} + \frac{\mu + \gamma}{\sigma} + \frac{1}{2}\sigma
\end{pmatrix}, \quad
d\mathbf{W}(t) := \begin{pmatrix} dW(t) \\ dW(t) \\ dW(t) \\ dW(t) \\ dW(t) \end{pmatrix}.
$$

Statistical Inference: Methodology

- (☞) Assume that the true value of the parameters $\beta_{s,0}$, $\beta_{a,0}$, p_0, and σ_0 are unknown.
- (☞) Use the quadratic variation of the SDE solution to estimate σ.
- (☞) Rewrite the SDEs by applying Lamperti's transformation to each component.
- (☞) Apply the Girsanov formula for the new system, and we obtained the likelihood function and the MLEs for $\beta_{s,0}, \beta_{a,0}, p_0$.

Estimator of the parameter of diffusion σ

$$\hat{\sigma}^2 := \sum_{i=1}^{5} \frac{\hat{\sigma}_i^2}{5},$$

where

$$\hat{\sigma}_1^2 = \frac{< S, S >_T}{\int_0^T (1 - S(t))^2 dt}, \qquad \hat{\sigma}_2^2 = \frac{< E, E >_T}{\int_0^T E(t)^2 dt}, \qquad \hat{\sigma}_3^2 = \frac{< I_a, I_a >_T}{\int_0^T I_a(t)^2 dt},$$

$$\hat{\sigma}_4^2 = \frac{< I_s, I_s >_T}{\int_0^T I_s(t)^2 dt}, \qquad \hat{\sigma}_5^2 = \frac{< R, R >_T}{\int_0^T R(t)^2 dt}.$$

MLEs for infection rates

$$\begin{pmatrix}\hat{\beta}_s \\ \hat{\beta}_a\end{pmatrix} = -\sigma \begin{pmatrix} J_s(T) & J_{sa}(T) \\ J_{sa}(T & J_a(T)\end{pmatrix}^{-1} \begin{pmatrix} \int_0^T \left[\frac{S(t)I_s(t)}{(1-S(t))} + \frac{S(t)I_s(t)}{E(t)}\right] dW(t) \\ \int_0^T \left[\frac{S(t)I_a(t)}{(1-S(t))} + \frac{S(t)I_a(t)}{E(t)}\right] dW(t)\end{pmatrix}.$$

whith

$$J_\star(T) := \int_0^T \left(\left[\frac{S(t)I_\star(t)}{(1-S(t))}\right]^2 + \left[\frac{S(t)I_\star(t)}{E(t)}\right]^2\right) dt,$$

$$J_{sa}(T) := \int_0^T I_a(t)I_s(t) \left(\left[\frac{S(t)}{(1-S(t))}\right]^2 + \left[\frac{S(t)}{E(t)}\right]^2\right) dt.$$

MLE for proportion of asymptomatic

$$\hat{p}_{ML} = \frac{\sigma}{J_2(T)} \int_0^T \left[-\frac{\kappa E(t)}{I_a(t)} + \frac{\kappa E(t)}{I_s(t)} \right] dW(t),$$

with

$$J_2(T) := \int_0^T \left[\frac{\kappa^2 E^2(t)}{I_s^2(t)} + \frac{\kappa^2 E^2(t)}{I_a^2(t)} \right] dt.$$

Hypotheses There exists $T_0 > 0$ such that for all $t \in [0, T_0]$

(❍) $\mathscr{R}_0^D > 1$ where

$$\mathscr{R}_0^D := \frac{p\kappa\beta_s}{(\mu+\kappa)(\mu+\alpha_s)} + \frac{(1-p)\kappa\beta_a}{(\mu+\kappa)(\mu+\alpha_a)}.$$

(❍) The initial condition X_0^+ and parameter configuration φ are such that $\Omega^* \neq \emptyset$ where

$$\begin{aligned}\Omega^* := \{(S, E, I_a, I_s, R) \times [t_0, T] : S(T) \leq S(t) < S(t_0), \\ E(t) > E(t_0),\ I_a(t) > I_a(t_0),\ I_s(t) > I_s(t_0)\}.\end{aligned}$$

Theorem

The estimators $(\hat{\beta}_{s,ML}, \hat{\beta}_{a,ML}, \hat{p}_{ML})$ are strongly consistent; that is,

$$\lim_{T \to T_0} \begin{pmatrix} \hat{\beta}_{s,ML} \\ \hat{\beta}_{a,ML} \\ \hat{p}_{ML} \end{pmatrix} = \begin{pmatrix} \beta_{s,0} \\ \beta_{a,0} \\ p_0 \end{pmatrix}, \qquad \text{with probability one.} \tag{1}$$

See [Baltazar-Larios et al., 2022]

Parameters of the simulation

Using the Milstein scheme, we divide the interval $[0, T]$ into n sub-intervals of length $\Delta = \frac{T}{n}$ and we obtain observations of each process at the times $0 = t_0 < t_1 < \ldots < t_n = T$ where $t_i - t_{i-1} = \Delta, i = 1, \ldots, n$.

We simulate a 1000 datasets in the time interval $[0, 1]$. We use $\beta_a = 0.251521$, $\beta_s = 0.456391$, $p = 0.1213$, $\sigma = 1/5000$, values of Table 1 and $\Delta = 1/850$.

Parameter	Real Value	Estimator	SD
β_s	0.456391	0.453155	2.900549E-005
β_a	0.251521	0.265229	6.081389E-002
p	0.1213	0.120354	2.554605E-006
σ	1/5000	1.361178E-004	5.325816E-009

Table: Average and Standard Deviation (SD) of estimator of σ and MLEs for β_a, β_s, and p.

Data description

* we use the dataset of new symptomatic and confirmed COVID-19 reported cases daily in Mexico City, considering that its population is $N = 26446435$.
* The dataset contains 47 records in the period March 10 to April 25, 2020.
* We use these records I_S^O to construct the observations, the process $I_S := I_S^O/N$.
* Combining I_S and the Milstein scheme, we generate the others four processes of the system with the parameters of Table 1, $\sigma = 1/100$ and $\Delta = 1/1000$.

Algorithm 1 Construction of dataset.

1: Using initial conditions and make $n = 0$.
2: Generate $\Delta W \sim N(0, \Delta)$.
3: $S(t_{n+1}) = S(t_n) - (\mu + \beta_a I_a(t_n) + \beta_s I_s^{mx}(t_n))\Delta S(t_n) + (\mu + \gamma R(t_n)) * \Delta - \sigma(1 - S(t_n))\Delta W$
4: $E(t_{n+1}) = E(t_n) - (\kappa + \mu) * \Delta E(t_n) + \Delta * (\beta_s I_s^{mx}(t_n) + \beta_a I_a(t_n) S(t_n) - \sigma(E(t_n))\Delta W$
5: $I_a(t_{n+1}) = I_a(t_n) - p\kappa E(t_n)\Delta I_a(t_n) - (\alpha_a + \mu) I_a(t_n)\Delta - \sigma I_a(t_n)\Delta$
6: $R(t_{n+1}) = 1 - S(t_{n+1} - E(t_{n+1}) - I_a(t_{n+1}) - I_s^{mx}(t_{n+1})$.
7: If $n < 47$ make $n = n + 1$ and go to 2. In other case stop.

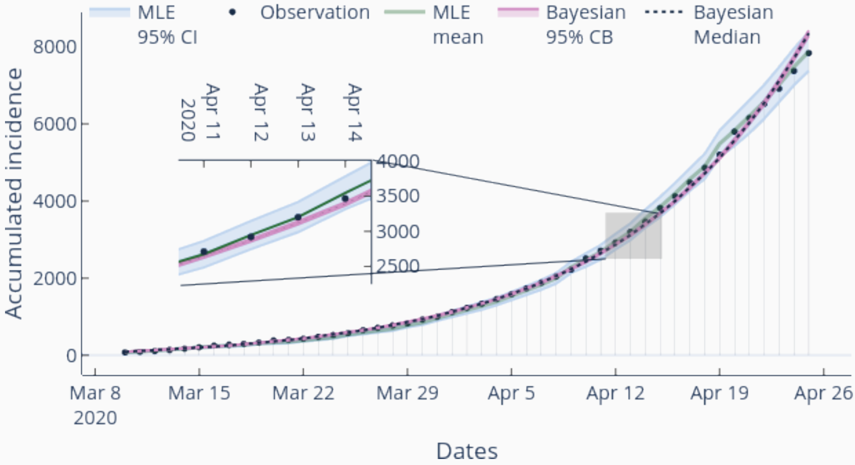
Estimators for real data

We assume the following parameters as given: $\gamma = 1/365$, $\kappa = 0.196078$, $\alpha_a = 0.167504$ $\alpha_s = 0.092507$. We generate 1000 datasets using the data of Mexico City.

Parameter	Estimator	CIL	CIU
β_s	0.059159	0.002546	0.115772
β_a	0.509925	0.378248	0.641603
p	0.582808	0.582326	0.583289
σ	1.31701E-002	8.32156E-003	1.94434E-002

Table: Average and 95% confidence interval of the estimator of σ and MLEs for β_a, β_s, and p.

Likening Between CDMX data fitting with MCMC and MLE.



Validation of the model

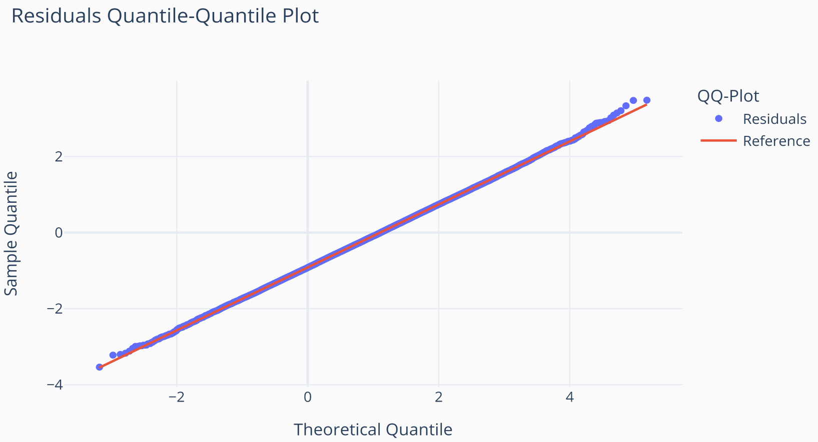


Figure: Quantile-Quantile plot of the numerical residuals.