In general the **posterior distribution** is *difficult* to obtain analytically, therefore **numerical integration** is required.

0.1 Monte Carlo approaches

In order to **compute the likelihood** as a starting point, if our knowledge is none, we can use a **grid of points** in order to compute $p(\vec{\theta}|y)$ for $\theta_1, \theta_2, ...\theta_n$ equally spaced. In this way, we *approximate* the continuous density function $p(\vec{\theta}|y)$ with the **discrete density function**:

$$\frac{p(\theta_i|y)}{\sum_i p(\theta_i|y)}$$

However, this approach is to become rapidly *unfeasibile* as the dimensionality of the parameter space becomes larger.

Another approximation one might want to take is the **Trapezoidal approximation**: after computing $p(\theta|y)$ for a discrete set of points $\theta_1, ...\theta_n$ we can approximate $p(\theta|y)$ with a piecewise-linear function, connecting the $p(\theta_i|y)$ points with linear segments.

One may ask now how to **numerically sample** $p(\theta|y)$. In order to do so, we use a *positive function* $g(\theta)$ defined for all θ such that $p(\theta|y) > 0$. Moreover it must hold that:

- we are able to draw samples from $g(\theta)$;

- for some *constant* M the ratio $p(\theta|y)/g(\theta) \leqslant M$ is defined;

- $g(\theta)$ must have a finite integral.

The **rejection sampling** algorithm hence is constituted by these two steps (Fig. 1):

- sampling θ at random from the pdf of $g(\theta)$;

- with probability that is given by the ratio $p(\theta|y)/Mg(\theta)$ we accept θ pretending it is drawn from the real $p(\theta|y)$ we wanted to approximate.
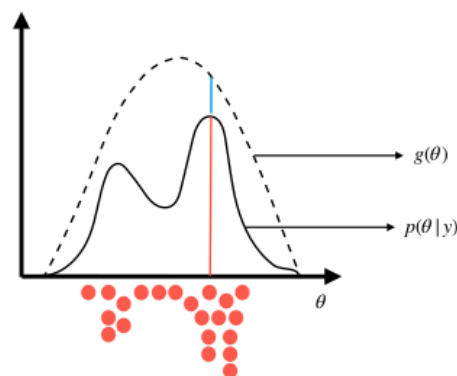
Figure 1: Rejection sampling method in order to sample from a distribution

The choice $g(\theta)$ plays an important role, specially wrt computational effort: the more different it is from $p(\theta|y)$, the higher is the rejection rate. However, if we have no information about the shape of $p(\theta|y)$ we should use a flat $g(\theta)$ (see Fig. 2a). On the other hand using the *trapezoidal approximation* to define $g(\theta)$ we can reduce this problem (see Fig. 2b).
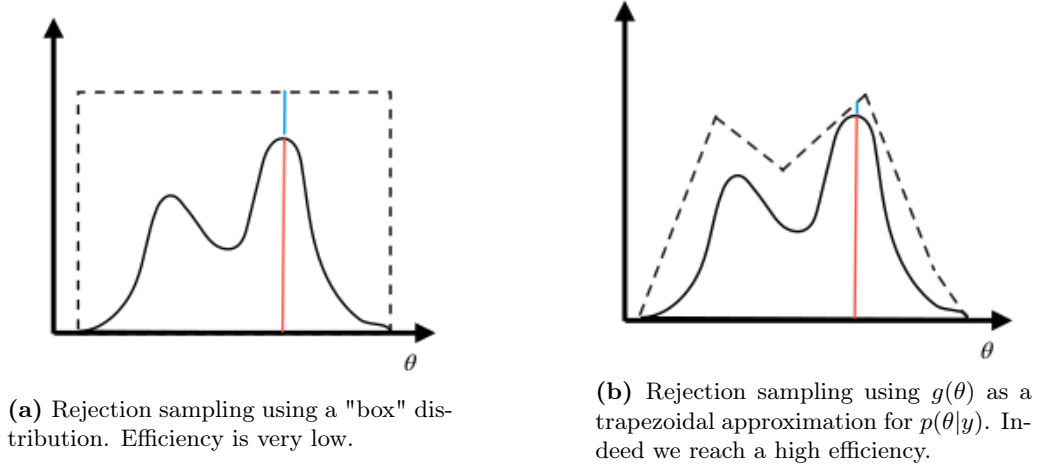
(a) Rejection sampling using a "box" distribution. Efficiency is very low.

(b) Rejection sampling using $g(\theta)$ as a trapezoidal approximation for $p(\theta|y)$. Indeed we reach a high efficiency.

Figure 2

We will now introduce the most efficient way to sample from distribution: **Markov Chain Monte Carlo** (MCMC). The general idea behind them is to start from an initial point in the parameter space θ_0, and starting from it to create a *random walk*. It is a *sequence* $\theta_0, \theta_1, ...\theta_t$ where each θ_i is draw from a given *transition distribution*, built such that the random walk converges to $p(\theta|y)$. That is to say we need to run the simulation long enough such that the distribution of the current draws becomes close enough to the stationary distribution $p(\theta|y)$. An example of MCMC is the so called **Metropolis algorithm** (see Fig. 3a, 3b). As in the general case we draw samples from a starting point θ_0 and then we iterate the following procedure for $t = 1, 2...$:

- sample a candidate point θ^* from a jumping distribution $J_t(\theta^*|\theta^{t-1})$. This distribution *must* be **symmetric**, i.e. such that $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a) \quad \forall a, b, t$;

- compute the ratio of the densities:

$$r = \frac{p(\theta^*|y)}{p(\theta^{t-1}|y)}$$

- update $\theta^t = \theta^*$ with probability $\min(r, 1)$ (Fig. 3b), otherwise $\theta^* = \theta^{t-1}$ (Fig. 3a).

Note that it is possible to show that for a *Metropolis Algorithm*, the Markov chain converges to a stationary distribution, which is $p(\theta|y)$. One may want to apply what we discussed so far and simulate *multiple chains* simultaneously with starting points dispersed in the parameter space. We keep monitoring the quantity (i.e. the parameter) of our interest and measuring variation between and within the different sequences, until all of them **converge** to a **unique distribution**. Therefore, the initial part of the random walk has to be neglected since samples are strongly correlated, we refer to this fact as **burn-in**.
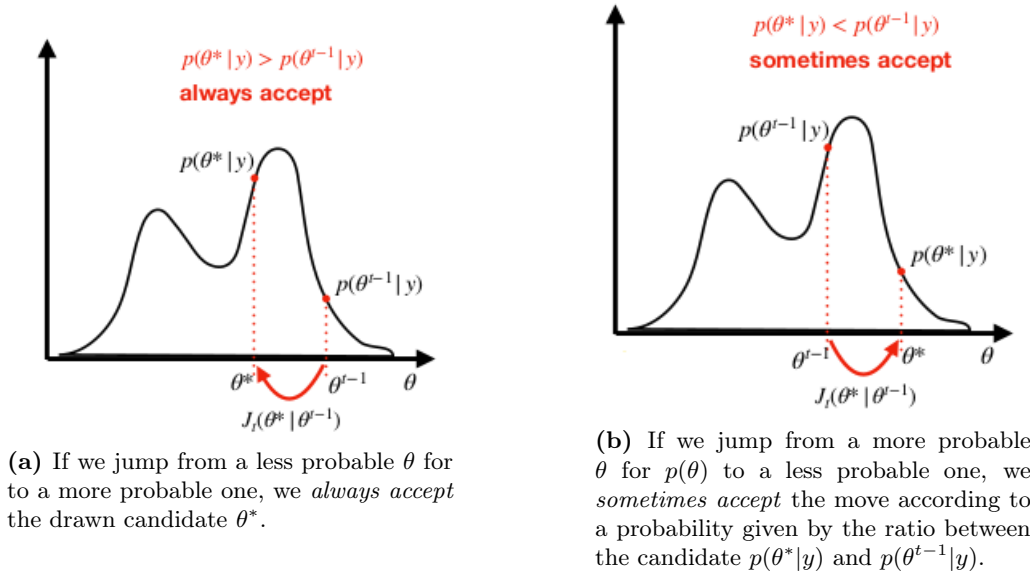
Let us now consider the following example: we want to sample from a posterior density that is a **bivariate unit normal**. The probability density function is therefore:

$$p(\theta_1, \theta_2|y) = \mathcal{N}(\theta_1, \theta_2|0, \mathbb{1}_2) = \frac{1}{2\pi}e^{-\frac{\theta_1^2+\theta_2^2}{2}}$$

where $\mathbb{1}_2$ is the 2×2 identity matrix.

As for the **jumping distribution** of the kind $J_t(\theta^*|\theta^{t-1}) = \mathcal{N}(\theta^*|\theta^{t-1}, \sigma^2\mathbb{1}_1)$, that we recall *must* be **symmetric**, we can choose a *bivariate normal* as well:

$$J(\theta_1^*, \theta_2^*|\theta_1^{t-1}, \theta_2^{t-1}) = \frac{1}{2\pi\sigma^2}e^{-\frac{(\theta_1^*-\theta_1^{t-1})^2+(\theta_2^*-\theta_2^{t-1})^2}{2\sigma^2}}$$

(a) If we jump from a less probable θ for to a more probable one, we *always accept* the drawn candidate θ^*.

(b) If we jump from a more probable θ for $p(\theta)$ to a less probable one, we *sometimes accept* the move according to a probability given by the ratio between the candidate $p(\theta^*|y)$ and $p(\theta^{t-1}|y)$.

Figure 3: Metropolis Algorithm.

The only **parameter** we need to tune as for the *jumping distribution* is σ:

- **large** σ: the larger it is, the larger jumps will be allowed such that we will be able to explore a larger space of parameters. On the other hand, almost all proposal will be rejected. Hence, despite the large inefficiency, we will end up being stuck;

- **small** σ: the chain will be too slow and the algorithm will be inefficient;

- **best** σ: we use as a *rule of thumb* a condition over the *acceptance rate* that has to be in the interval $0.23 - 0.44$.
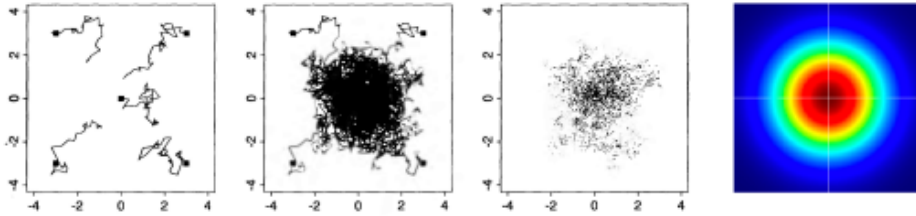

Figure 4: Markov Chain Monte Carlo at different stages of simulation. The distribution we use to sample from are bivariate Gaussians.

1

Outbreak Analysis

Let us apply all the knowledge we have acquired so far to a real world problem. We want to understand what we need to do when we observe an **outbreak** of an **emerging pathogen** about which we know *nothing*. Let us immerse ourselves in the following situation: to us all epidemiological characteristic, namely μ, β, generation time, are unknown. We will use COVID-19 as a *paradigmatic* case. For Fig. 1.1 one can tell that we achieved very good results in studying the disease as fast as possible: indeed all relevant quantities were computed in a couple of months.
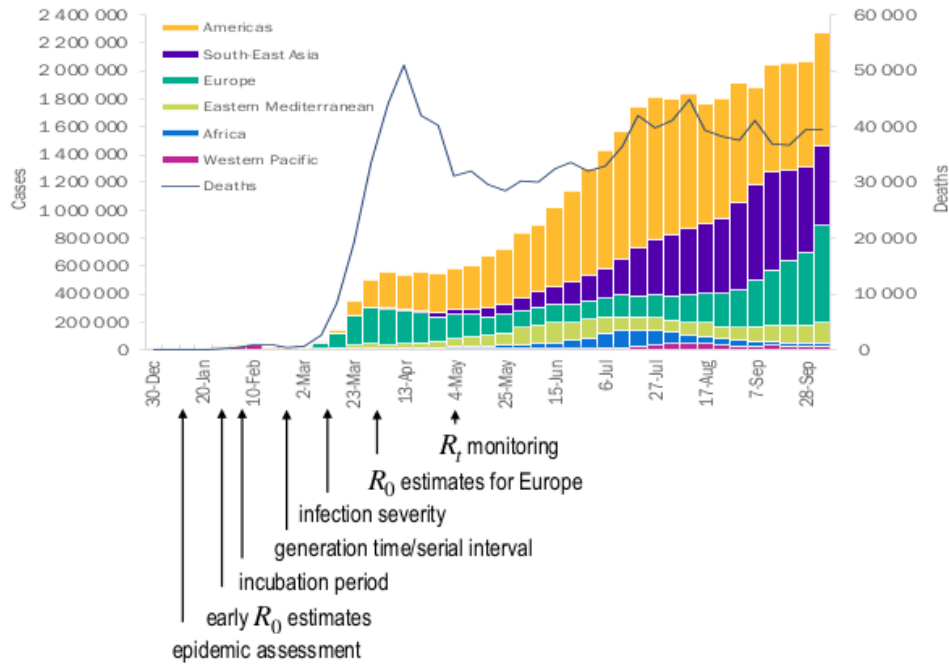
Figure 1.1: COVID-19 outbreak analysis Number of COVID-19 cases reported weekly by WHO Region, and global deaths, 30 December 2019. Arrows describe the moment at which the quantity reported has been computed. The epidemic assessment started from the letter written by Imperial College.

1.1   Basic Reproductive Ratio R_0 estimation

We will discuss now how to **process** the raw data in order to obtain, for instance, the **basic reproductive ratio** R_0. It is a really relevant quantity since we can retrieve the *final attack rate* if the virus spread: this is helpful in policy making to understand how much serious the disease can be, as well as how far we are from the epidemic threshold and how many infections/deaths we might expect.

1.1.1 R_0 from the early exponential growth

We will try to obtain R_0 using the *early exponential growth* typical of epidemic spread. This is indeed useful for **real time analyses** of highly transmissible infections (i.e. you see an exponential growth of cases).
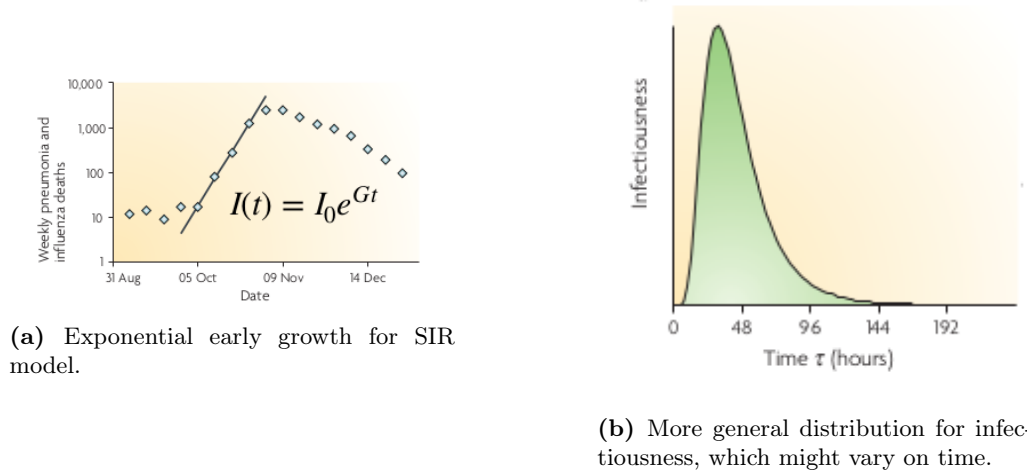
The estimate of R_0 will therefore computed by the means of the **exponential growth**, and obviously it is valid in the case we observe an epidemic. We will start to tackle this problem first introducing a simple version of the *Kermack and the McKendrick model*[1] according to which the **incidence** at the **early stages is** (Fig. 1.2a):

$$I(t) = I_0 e^{Gt} = I_0 e^{\mu(R_0 - 1)t} \tag{1.1}$$

We must observe indeed a region where cases grow exponentially: here we fit using the curve just introduced. Reverting this last formula we are able to **estimate R_0** from the *exponential growth factor G*:

$$R_0 = \mu^{-1} G + 1 \tag{1.2}$$

In this way we are **implicitly assuming** that *infectivity* β remains constant for the whole period of infection. It is indeed a strong, as well as unrealistic, assumption. Moreover, we assume that *recovery* is a Poisson process, whose waiting times are exponentially distributed with average μ^{-1}.

(a) Exponential early growth for SIR model.

(b) More general distribution for infectiousness, which might vary on time.

Figure 1.2

We define the **Generation time T_c** as the average time from infection of the infector to the infection of the infected. In this case, it holds that the average generation time is the average infectious duration $T_c = \mu^{-1}$, hence we can rewrite:

$$R_0 = T_c G + 1 \tag{1.3}$$

However, this means that **infectious periods** are **exponentially distributed**, which is quite *unrealistic*[2]. Another *unrealistic* aspect is the **lack** of **exposed period**, which is quite always present.

---

[1] The Kermack-McKendrick model is an SIR model for the number of people infected with a contagious illness in a closed population over time. It also assumes a completely homogeneous population with no age, spatial, or social structure.

[2] Wallinga, Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers Proc. R. Soc B (2007) 274, 599

Let us consider the more general situation as possible where the *generation times* distribution is $w(\tau)$. We recall now that the **transmissibility generic function** $\beta(\tau)$ can be rewritten as function of the generating time as:

$$w(\tau) = \frac{\beta(\tau)}{R_0} \tag{1.4}$$

This means that there is a period in which we are more likely to infect someone once infected, i.e. we are more infectious. Obviously, this is the peak in the distribution $w(\tau)$ (see Fig. 1.2b). The other way around to see it is the following: the time needed in order to generate a secondary case, which might be in number more than one, distributes as in Fig. 1.2b.

Let us imagine to have a number $I(u)$ of newly infected individuals at time u. The **number of individuals they will infect** at time $t = u + \tau$ in a time interval $\delta\tau$ is a random process, and follows a **Poisson** distribution with mean:

$$I(u)\beta(\tau)\delta\tau = I(t-\tau)\beta(\tau)\delta\tau \tag{1.5}$$

One can easily see that the number of **newly generated** infected people *does* depend on the number of *previously infected* individuals. Formally, this is a **renewal equation** (Lotka-Euler equation)[3]:

$$I(t) = \int_0^\infty I(t-\tau)\beta(\tau)d\tau \tag{1.6}$$

Simple SIR model actually accounts for these last considerations, but it assumes that generation times distribution is exponential. We wanted to drop this assumption: we are indeed considering non-markovian epidemics. For this reason we had to introduce the aforementioned **integral equation**, which takes into account the exponential case too, and we could not write an exponential equation any more. Only for now, we are assuming well-mixed and therefore infinite population.

We want now to use the *renewal equation* in order to estimate R_0. This can be done however *only* when we observe an **exponential growth** of cases, according to which we can derive the *exponential growing factor G*. Hence we can write the infectious people at time $I(t)$ in function of infectious people at time $t - \tau$ as follows:

$$I(t) = I(t-\tau)e^{G\tau} \tag{1.7}$$

By using last equality, the renewal equation becomes:

$$I(t) = \int_0^\infty I(t)e^{-G\tau}\beta(\tau)d\tau \tag{1.8}$$

Dividing both sides by $I(t)$ and replacing 1.4 in the last equation, we end up with:

$$\frac{1}{R_0} = \int_0^\infty e^{-G\tau}w(\tau)d\tau \tag{1.9}$$

From which we can provide an estimate for R_0. One should have noted that it is the **Laplace Transform** of $w(\tau)$: therefore we are able to relate directly R_0 to the generation times $w(\tau)$ distribution, specially to its moments generating function via $\mathcal{M}(z) = \int_0^\infty e^{z\tau}w(\tau)d\tau$:

$$R_0 = \frac{1}{\mathcal{M}(-G)} \tag{1.10}$$

We recall now the **properties** of the *Laplace Transform* and its **epidemiological implications**:

---

[3]a.k.a., borrowed from population dynamics. New births depend on people that were born 20-30 years ago: for instance their number (larger it is, the more births we expect), fecundity rate and how this parameter evolves during the whole lifespan of a person. Many models from ecology and demography can be applied to the epidemiological framework.

- $\mathcal{M}_{w(\tau)}(z) \geqslant 0 \implies R_0 \geqslant 0$ because the Laplace transform is *always* greater or equal than zero.

- $\mathcal{M}'_{w(\tau)}(z) = \int_0^\infty \tau e^{z\tau} w(\tau) d\tau > 0 \implies R_0$ is an increasing function of G, since the derivative of the Laplace transform is always greater than zero. Bigger the exponential growth, bigger is R_0.

- $\mathcal{M}_{w(\tau)}(0) = \int_0^\infty w(\tau) d\tau = 1 \implies$ if $G = 0 \implies R_0 = 1$ from normalization condition. If we observe a flat curve, i.e. no exponential growth, then $R_0 = 1$.

Let us assume that $w(\tau)$ is distributed as a Gaussian, that we recall is the least informative distribution[4]. We know from theory that its *Laplace Transform* is:

$$w(\tau) = \mathcal{N}(T_c, \sigma^2) \implies R_0 = e^{GT_c - (1/2)G^2\sigma^2}$$

This means that R_0 depends also on the variance of generation times distribution. Let is assume that the generation time is like the infectious period ($\langle T_c \rangle \sim \mu^{-1}$). For a given $\langle T_c \rangle$, if the variability σ^2 of the generation times is higher, then the exponential term becomes lower, hence leading to lower R_0 (see Fig. 1.3). For instance: if an individual remains infected for a shorter period, it is going to produce quicker more infections hence the exponential growth will become steeper. This fast chain will dominate over the other chain of transmissions.

(a)

(b)
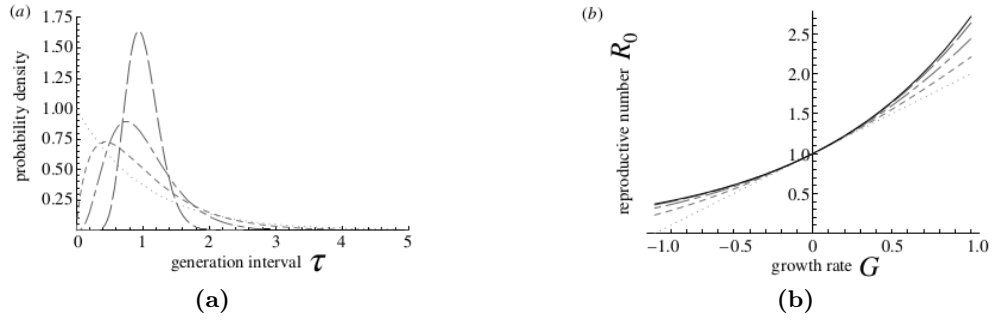
Figure 1.3: Different distributions to of generation times (1.3a) lead to different R_0 once we observe an exponential growth G (1.3b).

Whereas, if we consider a **Delta Distribution**, i.e. a well peaked distribution with no variance at all, its transform is:

$$w(\tau) = \delta(T_c) \implies R_0 = e^{GT_c}$$

that sets an **Upper bound** of R_0. This is the case where the infection duration is more or less equal for everybody.

In other words, adding some variability onto a fixed $\langle T_c \rangle$ generation time, a given R_0 will correspond to higher G. This can be formalized thanks to **Jensen's inequality**: the average of transformed stochastic variables is at least equal to the transformed average of those variables when the transformation is *convex*:

$$\int_0^\infty e^{z\tau} w(\tau) d\tau \geqslant e^{z \int_0^\infty \tau w(\tau) d\tau} \implies \mathcal{M}_{w(\tau)}(z) \geqslant e^{zT_c} \implies R_0 \leqslant e^{GT_c} \qquad (1.11)$$

Recalling that τ is distributed according to some distribution $w(\tau)$, we assume that τ is the sum of two stochastic variables:

$$\tau = \tau_E + \tau_I$$

---

[4]It means that we know only the mean μ and the variance σ^2 of a distribution and nothing more.

where τ_E is distributed according to $g(\tau_E)$ and τ_I follows another distribution, namely $h(\tau_I)$. One can prove that the generation times distribution $w(\tau)$ can be rewritten as the *convolution* between g and h:

$$w(\tau) = g(\tau_E) * h(\tau_I) \tag{1.12}$$

and, its Laplace transform is:

$$\mathcal{M}_{w(\tau)} = \mathcal{M}_{g(\tau_E)} \times \mathcal{M}_{h(\tau_I)} \tag{1.13}$$

This can turn out to be useful if we want to add some complexity to our model[5] and consider the generation time as the sum of a **latent period**, during which *no infections* can be generated and that distributes as $g(\tau_E)$, and a second term that is an **infection period** τ_I where infections can be generated and distributes as $h(\tau_I)$. For instance, let as assume that $\tau_E \sim Exp(\varepsilon)$ $\tau_I \sim Exp(\mu)$. The basic reproductive ratio is defined as $R_0 = \frac{1}{\mathcal{M}(-\mathcal{G})}$: let us compute it analytically. We know that:

$$\frac{1}{R_0} = \mathcal{M}(-G) = \int_0^\infty e^{-G\tau}\omega(\tau)d\tau \tag{1.14}$$

Given $g(\tau_E) = \varepsilon e^{-\varepsilon\tau_E}$ and $g(\tau_I) = \mu e^{-\mu\tau_I}$

$$\mathcal{M}(-G) = \mathcal{M}_{g(\tau_E)} \times \mathcal{M}_{h(\tau_I)} = \int_0^\infty e^{-G\tau_E} g(\tau_E)d\tau_E \times \int_0^\infty e^{-G\tau_I} g(\tau_I)d\tau_I$$

$$= \int_0^\infty e^{-G\tau_E}\varepsilon e^{-\varepsilon\tau_E}d\tau_E \times \int_0^\infty e^{-G\tau_I}\mu e^{-\mu\tau_I}d\tau_I$$

$$= \varepsilon\int_0^\infty e^{-(G+\varepsilon)\tau_E}d\tau_E \times \mu\int_0^\infty e^{-(G+\mu)\tau_I}d\tau_I$$

$$= \frac{\varepsilon}{G+\varepsilon} \times \frac{\mu}{G+\mu} = \frac{1}{R_0}$$

The **basic reproductive ratio**:

$$R_0 = \left(1 + \frac{G}{\varepsilon}\right) \cdot \left(1 + \frac{G}{\mu}\right) = 1 + \frac{G}{\varepsilon} + \frac{G}{\mu} + \frac{G^2}{\varepsilon \cdot \mu} \cong 1 + G(\varepsilon^{-1} + \mu^{-1}) \tag{1.15}$$

One should note that we have exactly got back to the result of the **SEIR** model.

One should note however that if we observe an **exponential growth**, we cannot conclude *anything* on the *Basic Reproductive Ratio* R_0, since we need both average and variance information of **generation time distribution** as well: this indeed helps to interpret the exponential growth (Fig. 1.3b).

In practice, when we have an **incidence time series** $\{y_t\}$ we need to:

- *Define a time window* where the **growth** is **exponential**.

- We make a **Poisson regression** of the incidence points, assuming that our process is Poisson and with expected value:

$$\log\left(\mathbb{E}(y_t|t)\right) = \alpha_0 + Gt$$

- We estimate through the *likelihood* both **initial conditions** and the **exponential growth**.

$$\mathcal{L}(\alpha_0, G|\{y_t\}) = \prod_{t=0}^{t_M} \frac{e^{y_t(\alpha_0 + Gt)}e^{-e^{\alpha_0 + Gt}}}{y_t!}$$

---

[5]What we observe in reality is that infectivity continuously evolves throughout time: sometimes it is less and sometimes is high. We can simplify and divide this behavior into two discrete stages: Exposed and Infected.

Initial conditions are fundamental, since we need them to properly understand and interpret G. Even though we want to infer a single parameter and α_0 has no specific meaning taken by itself, those are needed.

Clearly, the proper distribution of generation times is difficult to obtain analytically; therefore, we use only available information on the generation time distribution and make some assumptions in order to compute R_0. Many times the only information we have is its *average T_c* and *dispersion σ*, hence we can exploit the **Gaussian approximation**:

$$R_0 \simeq e^{GT_c - (1/2)G^2\sigma^2}$$

However, often the **distribution of serial intervals** is used as a *proxy* of the distribution of generation time.