Mathematical Modeling of Infectious Diseases, UCSF

Version 1.0—modified 10 Mar 2016

Markov and Hidden Markov models

We've spent time now on classical compartmental models. We will now return to our stochastic roots so to speak. Instead of a continuous number of people of each type, we will have an integer count that changes, jumping up or down when something happens.

But first a quick review of a few things.

Let's warm up with a simple example. Remember before we considered models with a constant arrival? Deterministically, we would have

$$\frac{dX}{dt} = \Lambda$$

Of course this is duller than ditchwater; the solution is just $X(t) = \Lambda t + X(0)$. You have however many people you started with, plus the arrival rate times the elapsed time. In point of fact we saw this embedded in the so-called immigration death process, whose simple deterministic form was

$$\frac{dX}{dt} = \Lambda - \mu X$$

Here, arrival is constant, but death is constant per capita. If X is huge, deaths outweigh the feeble constant arrival and the population shrinks; if X is tiny, arrivals outweigh the few deaths and the population grows. We saw that the equilibrium value, where inflow balances outflow, occurs at the value $\bar{X} = \Lambda/\mu$.

Although this equation has a simple analytic solution, it's interesting to imagine that $X = \bar{X} + \epsilon$. We suppose at the outset that we perturb away from the equilibrium by a tiny amount ϵ. Note

$$\frac{dX}{dt} = \frac{d\epsilon}{dt}$$

and substituting

$$\frac{d\epsilon}{dt} = \Lambda - \mu(\bar{X} + \epsilon)$$
$$\frac{d\epsilon}{dt} = \Lambda - \mu\bar{X} - \mu\epsilon$$
$$\frac{d\epsilon}{dt} = \Lambda - \mu\frac{\Lambda}{\mu} - \mu\epsilon = -\mu\epsilon$$

If $\epsilon > 0$, its growth rate is negative—a small positive perturbation gets smaller. If $\epsilon < 0$, its growth rate is positive, so that it again takes us back toward the equilibrium.

This technique is much more important in nonlinear systems, where it is called *local stability analysis*. Time doesn't permit us to say much more about it, but it's a fundamental tool in exploring compartmental models in epidemiology as well as many other things.

Exercise 1. Write the classical SIR model using an incidence term of the form βXY and of the form $\beta' X\frac{Y}{N}$. What is the R_0 in each case? How does it depend on population size in each case? Under what conditions is there a critical community size for disease persistence? (Optional part) Find a form for the incidence rate that interpolates between the two cases I gave you. ∎

Now, let's take another look at the simple inflow process. Again, deterministically it was just $\dot{X} = \Lambda$. Stochastically, what is a natural analog?

Suppose that X now is *integer*-valued, on 0, 1, 2, and so on. This counts the number of people who have arrived. Suppose each infinitesimal unit of time dt, the chance of an arrival is Λdt, constant. Each tiny dt is like any other; we have a constant arrival rate. We know the waiting time till the next arrival must be exponential, based on theory we've already done. When the next person arrives, the count will jump. We could easily simulate this process by drawing random exponentials and using them to find the jump times.

Let $p_i(t)$ be the probability that X is i at time t. It is the chance there have been i arrivals. At the outset, $p_0(0) = 1$ and $p_i(0) = 0$ for $i > 0$; this means no one has arrived at the start. We will use the law of total probability to write the motion:

$$p_i(t + dt) = \sum_{j=0}^{\infty} p_j(t) P(X_i(t + dt) = i | X_i(t) = j)$$

It's just this: the chance you were at someplace j a moment ago, times the chance you would have come to j from i during the time interval, added up for all possible places you could have been.

By definition, you can't lose anybody, so if $i < j$ forget it; the transition probability is 0.

If there is a chance of one arrival of Λdt during the interval of length dt, then two arrivals would have chance $\Lambda^2(dt)^2$. We usually just write $o(dt)$ to mean something that goes to zero as $dt \to 0$. Basically all that can happen in an infinitesimal dt is you stay where you are at X, or you add 1.

The chance you add 1 is Λdt; the chance you don't add one must be $1 - \Lambda dt$. So

$$p_i(t + dt) = p_i(t)(1 - \Lambda dt) + p_{i-1}\Lambda dt$$

for $i > 0$ and

$$p_0(t + dt) = p_0(t)(1 - \Lambda dt)$$

for $i = 0$.

It's mildly irritating to have these two equations; we could get rid of one by just stating $p_{-1} = 0$ always, or by writing

$$p_i(t + dt) = p_i(t)(1 - \Lambda dt) + (1 - \delta_{i0})p_{i-1}\Lambda dt$$

using the symbol δ_{jk}. This sort of thing is called a Kronecker delta; it is 1 whenever $j = k$ and 0 otherwise. Maybe it has nobler uses in mathematics, but it won't hurt us to use it.

Let's work this equation over a bit.

$$p_i(t + dt) = p_i(t) - p_i(t)\Lambda dt + (1 - \delta_{i0})p_{i-1}\Lambda dt$$

$$p_i(t + dt) - p_i(t) = -p_i(t)\Lambda dt + (1 - \delta_{i0})p_{i-1}\Lambda dt$$

$$\frac{p_i(t + dt) - p_i(t)}{dt} = -p_i(t)\Lambda + (1 - \delta_{i0})p_{i-1}\Lambda$$

$$\frac{p_i(t + dt) - p_i(t)}{dt} = -p_i(t)\Lambda + (1 - \delta_{i0})p_{i-1}\Lambda$$

Then of course take the limit

$$\frac{dp_i(t)}{dt} = -p_i(t)\Lambda + (1-\delta_{i0})p_{i-1}\Lambda$$

For $i = 0$ we just have $dp_0/dt = -\Lambda p_0$. So $p_0(t) = e^{-\Lambda t}$. The chance there are no people just decays away exponentially.

This set of equations is called the Kolmogorov forward equations for the process.

Exercise 2. Show that $p_i(t) = \frac{e^{-\Lambda t}(\Lambda t)^i}{i!}$ is the solution. Do this by dispensing first with $i = 0$; we already have that answer. Then just substitute this into the differential equations and show you get an identity that is always true. ∎

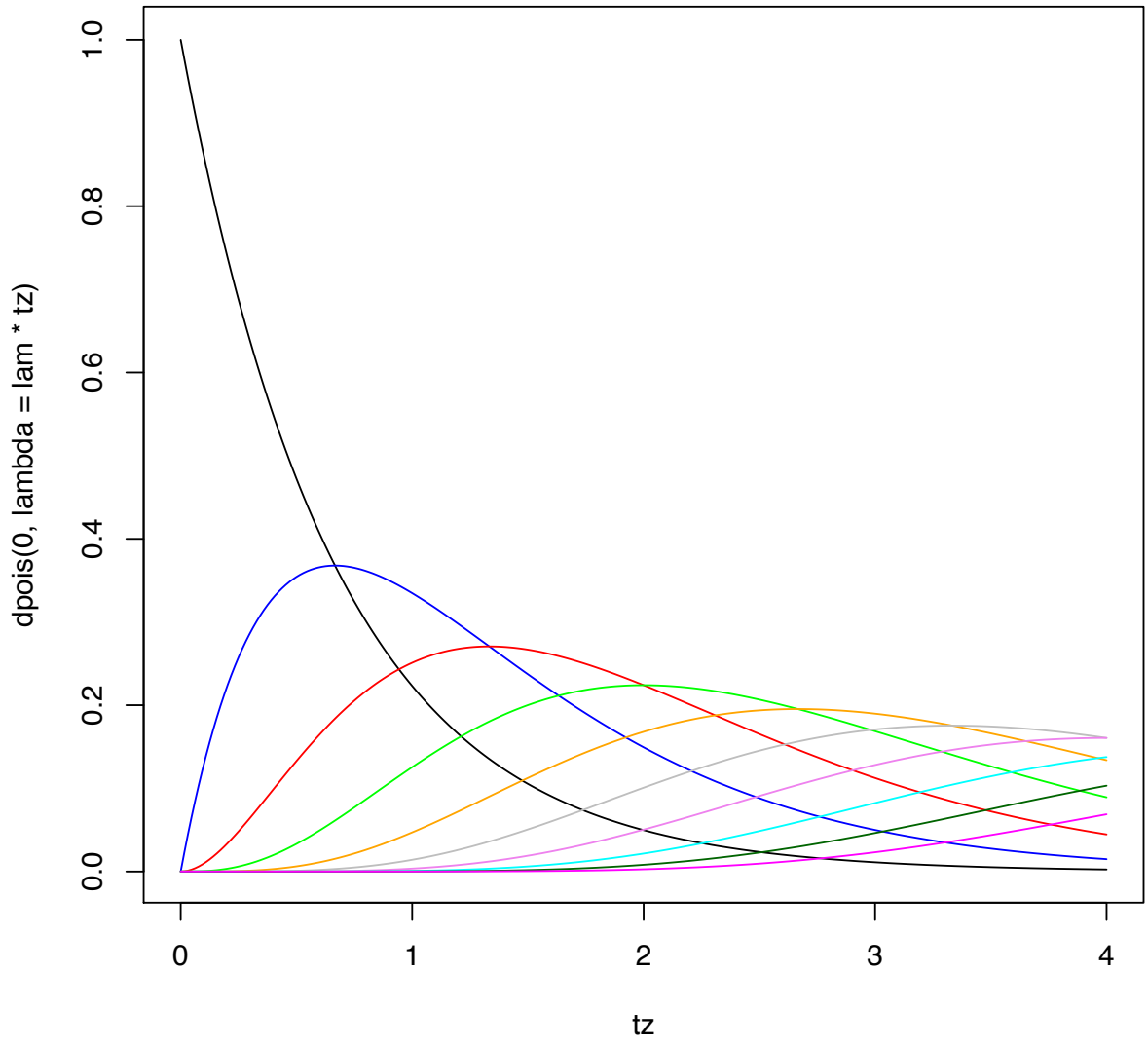
What the exercise shows you is that the count of arrivals between 0 and some time t is a *Poisson random variable*, whose mean is Λt. This is called a Poisson arrival process. You know the interarrival times are exponential, and the count in some window is Poisson with mean given by the rate of arrival times the length of the window. This process is a fundamental building block of modeling in many areas of operations research and not just population modeling.

What we have found is a time varying probability distribution. For small times, at the beginning, the mean is nearly zero. Later on, the mean becomes larger. Over time, the probability of some outcomes decreases, and that of others increases. Don't lose track: sometimes we call the mean of the Poisson distribution λ, but when talking about the Poisson process (like we're doing now), we call the rate λ, and the mean counts in $[0, t)$ is λt.

Let's take a look at what the probabilities look like over time. I want to assume $\lambda = 1.5$ arrivals per time, and I'll follow it for 4 units. Maybe these are patient arrivals per hour at a clinic for instance, and we're simulating the morning shift of 4 hours.

```
lam <- 1.5
tz <- seq(0,4,by=1/64)
plot(tz,dpois(0,lambda=lam*tz),type="n")
clrz <- c("black","blue","red","green","orange","gray","violet","cyan","darkgreen","ma
```

```
for (ii in 0:9) {
  points(tz,dpois(ii,lambda=lam*tz),type="l",col=clrz[ii+1])
}
```



Let's look at some realizations of the process, starting with a few. I'm going to code it

in a flexible simple way, but this is not really the fastest way in R.

```
ppsim <- function(lam,mxt) {
  ppsim.r(lam,0,mxt,c())
}
ppsim.r <- function(lam,curt,mxt,past) {
  new <- rexp(1,rate=lam)
  newt <- new + curt
  if (newt > mxt) {
    past
  } else {
    ppsim.r(lam,newt,mxt,c(past,newt))
  }
}
```

Let's run one:

```
ppsim(1.5,4)

## [1] 0.4725239 0.9823427 2.0055193 3.2018064 3.5511544 3.9535208
```

Those are the arrival times. Notice all we need to do is simulate exponential waiting times to find out where the jumps happen. Let's do some plots:

```
pp.plot <- function(lam,mxt,nsims=10,colrz=clrz) {
  mx <- 0
  l1 <- list()
  for (ii in 1:nsims) {
    l1[[ii]] <- ppsim(lam,mxt)
    mx <- max(mx,length(l1[[ii]]))
```

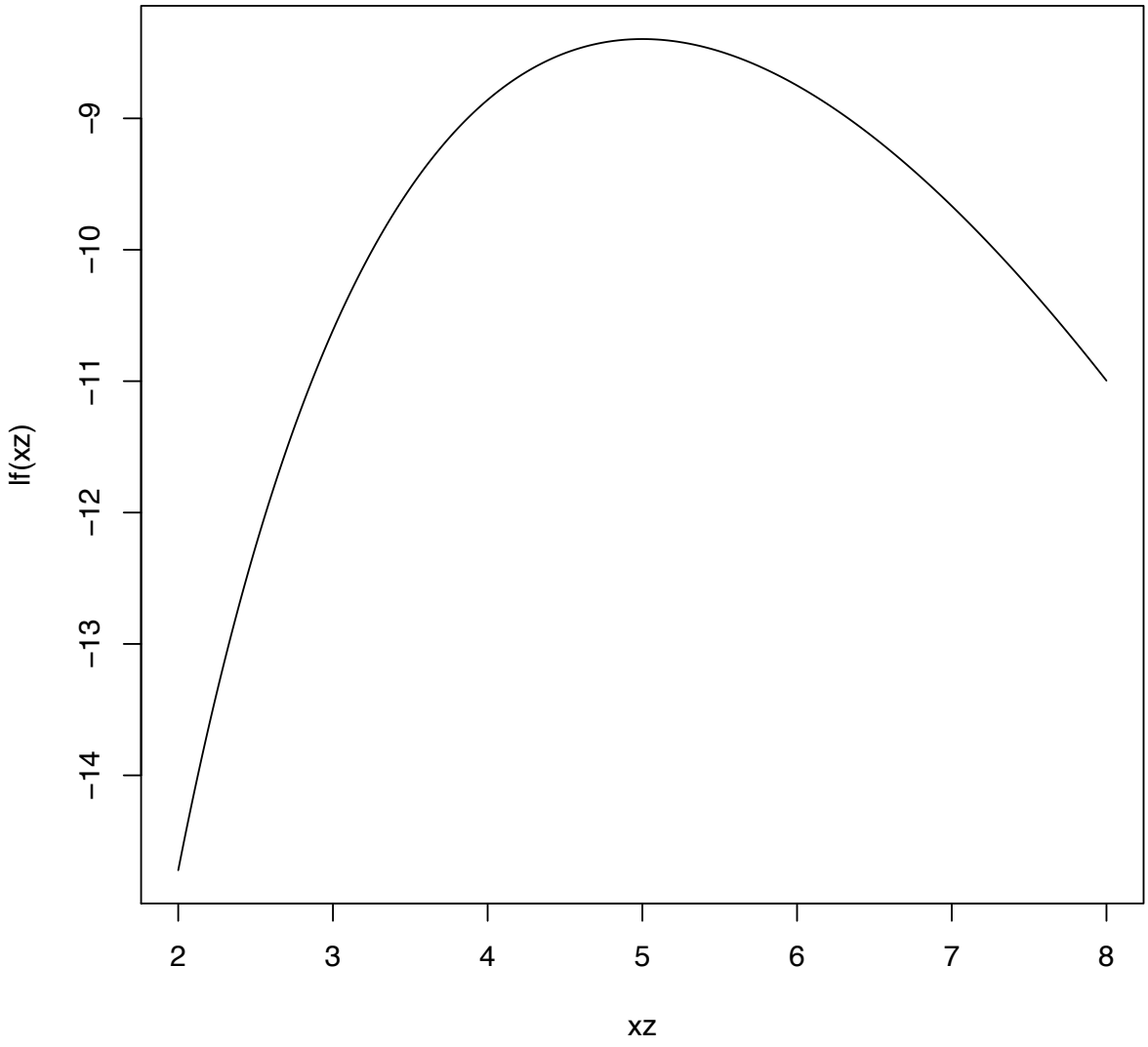
```
  }
  xz <- seq(0,mxt,length=256)
  plot(xz,xz,ylim=c(0,mx+1),xlab="Time",ylab="Count",type="n")
  for (ii in 1:nsims) {
    yz <- sapply(xz,function(x)sum(l1[[ii]]<x))
    points(xz,yz,type="l",col=colrz[ii])
  }
}
```

If we pick a time, and ask how many have arrived by that time, it's Poisson. As you do more and more realizations, the relative frequencies should look more and more like the probabilities.

Suppose we believe that $\lambda = 1.5$ and we observe 4 realizations of the process at time 4 say. For instance, we may have the total count of arrivals in 4 hours. Perhaps this is the counts on Monday, Tuesday, Wednesday, and Thursday. If our model is that these are independent and identically distributed, we can compute the probability of the data. Let's call the counts Y_1, Y_2, Y_3, and Y_4. Since the mean is $\mu = \lambda t = 6$, we can write $P(Y_1 = y_1) = e^{-\mu}\mu^{y_1}/y_1!$ for the first one, and so on for the others. We're just computing the probability of the data given the model as a function of the parameters, which is called the likelihood. Here, the likelihood depends on μ and thus on the rate λ. We assumed $\lambda = 1.5$, but really we might want to change λ and find the value that makes the data most likely. Say that the observations were 5, 7, 2, and 6. Let's plot the likelihood (or to make life simpler, the log of the likelihood):

```
lf <- function(lm) {
  tmp <- rep(NA,length(lm))
  for (ii in 1:length(lm)) {
    tmp[ii] <- sum(dpois(c(5,7,2,6),lambda=lm[ii],log=TRUE))
  }
  tmp
```

```
}
xz <- seq(2,8,by=1/128)
plot(xz,lf(xz),type="l")
```



```r
mean(c(5,7,2,6))

## [1] 5
```

You can see that the likelihood happens to peak at the sample mean in this simple problem.

So we looked at the Poisson arrival process.

Now, let's ratchet things up a notch. What if the individuals can die? Suppose there is a constant death rate too, per capita. As before, let X be the count; $p_i(t)$ is $P(X(t) = i)$ as before. We'll still have the same idea. Just now, in an infinitesmial time interval we could stay the same, add one, or lose one. The chance you gain one is Λdt. The chance you lose one is $i\mu dt$. The chance two or more things happen is $o(dt)$. We could write the chance we don't gain one is $1 - \Lambda dt$, and the chance we don't lose one is $1 - i\mu dt$, and the chance we neither gain nor lose is $(1 - \Lambda dt)(1 - i\mu dt) = 1 - \Lambda dt - i\mu dt$ to first order. We're saying that for infinitesimal dt, you gain one with chance Λdt, lose one with chance $i\mu dt$, and neither with chance $1 - \Lambda dt - i\mu dt$.

$$p_i(t + dt) = p_i(t)(1 - \Lambda dt - i\mu dt) + (1 - \delta_{0i})p_{i-1}\Lambda dt + p_{i+1}(i + 1)\mu dt$$

which will give us

$$\frac{dp_i(t)}{dt} = -p_i(t)(\Lambda + i\mu) + (1 - \delta_{0i})p_{i-1}\Lambda + p_{i+1}i\mu$$

This is the Kolmogorov equation for the immigration-death process in continuous time. Solving this is not going to be so simple, even though it is a set of linear equations with constant coefficients. It is an infinite set of them, and there are tricks to it. But for us, it's just a stepping stone to the next process we want to talk about.

That process is the SIS model. Now, imagine we have a population of susceptibles and infectives; we have N total people. We want to model the number of infectives. Assuming a random recovery rate, we could treat recovery the way we treated mortality: let $\gamma = \mu$. So let's start with

$$\frac{dp_i(t)}{dt} = -p_i(t)(\Lambda + i\gamma) + (1 - \delta_{0i})p_{i-1}\Lambda + p_{i+1}(i + 1)\gamma$$

But instead of a *fixed* inflow Λ, suppose that the inflow rate is proportional to the number of *susceptibles*. Let $\Lambda = \lambda(N - i)$.

$$\frac{dp_i(t)}{dt} = -p_i(t)(\lambda(N-i) + i\gamma) + (1 - \delta_{0i})p_{i-1}\lambda(N - i + 1) + p_{i+1}(i+1)\gamma$$

Why $N - i + 1$ in the second term? Because if there are $i - 1$ infectives, there must be $N - (i - 1)$ susceptibles.

Then, let's let $\lambda = \beta\frac{i}{N-1}$; we're just assuming that the hazard or force of infection is proportional to the prevalence fraction in the population as we have already done. It's $N-1$ in the denominator, since there are $N - 1$ other people in the population. It makes no difference really, since in this model N is constant anyway. So:

$$\frac{dp_i(t)}{dt} = -p_i(t)(\beta\frac{i}{N-1}(N-i) + i\gamma) + (1 - \delta_{0i})p_{i-1}\beta\frac{i-1}{N-1}(N - i + 1) + p_{i+1}(i+1)\gamma$$

So there you have it. This is the Kolmogorov forward equation for the SIS process. Written this way, no probability will ever get up to $N + 1$, since the flow rate from the state with $i = N$ up to $i = N + 1$ will be zero.

This whole thing is actually $N + 1$ *linear* equations for the probability of there being various numbers of infectives. Over time, infectives come and go, and the system as a whole follows these equations. We're going to numerically explore them. It turns out that you can actually use this model for small villages or communities as we have in the trachoma field.

Before going on, you might think: this is an epidemic model, and haven't we heard that it is supposed to be nonlinear? What about the bilinear incidence term? In the deterministic version, the model is indeed nonlinear, but the stochastic version is still a large linear set of equations. The bilinear incidence term appears as $i(N - i)$ for example—but i is a function of the index, not a dynamic variable. The dynamic variables are the probabilities; we have different coefficients for each equation, involving products of i and $N-i$ and whatnot, but the dynamic variables are p_i, and these never get multiplied by each other or anything remotely nonlinear. The equations are really linear! Moreover, they are actually a compartmental model in a way, but a compartmental model of probability; it's almost as though probability is some "stuff" flowing from place to place.

So, we have a bunch of differential equations for the chance the village is in a certain state, by which we mean really has a certain number of infectives. If we start at some time, say 0, we can just run this model, i.e. numerically calculate the solution of the equation in the future, at future times t. This takes us from the chance the system is in some state at one time to the chance it is in whatever state at a later time. (In fact you could precompute the values over some specific time period and have a big matrix saying the chance the system is in state i at time t given that it was in state j at time $t - t'$, a time t' ago. We won't pursue this.)

The point is that if you think of I as a random variable taking values from 0 anywhere up to N, this random quantity evolves over time. The equations tell us how to get from the distribution at one time to the distribution later.

We will numerically explore this model.

We need the hypergeometric distribution.

This is a continuous time Markov chain. Suppose at regular intervals we observe the process, say like doing a trachoma survey every 6 months. We don't actually know the number of infected people. We might have only imperfect lab tests, for instance. Or we might only sample the population randomly. Let I_0 be the value at the beginning, I_1 at the first observation time, and so on. We could write the DAG $I_0 \to I_1 \to I_2 \cdots$. But it's not enough; we would add arrows to it from $I_0 \to Y_0$, $I_1 \to Y_1$, and so on. In this model, our observations are conditionally independent of the nondescendants given the parents; each Y depends on the underlying state. You could imagine more complicated models where you would do it differently, but this is the assumption here.

This takes us to the world of the hidden Markov model. Using what we've learned about DAGs, we would write the full joint distribution as

$$P(I_0)P(I_1|I_0)P(I_2|I_1)\cdots P(Y_0|I_0)P(Y_1|I_1)\cdots.$$

Then, however, we need to sum over all values of the unobserved true (hidden) variables to

get the probability of the data:

$$\sum_{I_0, I_1, cdots} P(I_0)P(I_1|I_0)P(I_2|I_1)\cdots P(Y_0|I_0)P(Y_1|I_1)\cdots.$$

This is just the chance of the data. Here, $P(I_0)$ is the initial condition, probably involving parameters; each quantity of the form $P(I_{t+1}|I_t)$ is derived from the differential equations, and will involve β and γ, transmission and recovery rates. Let $A_{ji} = P(I_{t+1} = j|I_t = i)$.

What about $P(Y_0|I_0)$ and similar terms? These are the so-called emission probabilities.

Suppose that Y_t is the number of positive samples out of a random sample of size S. If there were really i infectives, then we would have

$$P(Y_t = y_t|I_t = i) = \frac{\binom{i}{y_t}\binom{N-i}{S-y_t}}{\binom{N}{S}}$$

if we assume random sampling. Of course if $y_t > i$ this would be zero for example.

Exercise 3. How long did this assignment take you? ∎