Mathematical Modeling of Infectious Diseases, UCSF

Homework 3

Version 1.1—modified 23 Jan 2015

Exercise 1. Prove $\text{cov}(X,X) = \text{var}(X)$. Prove $\text{cov}(X,aY) = a\text{cov}(X,Y)$ for a any constant. Prove $\text{cov}(X,Y) = \text{cov}(Y,X)$ for any random variables X and Y (as long as $E[X]$, $E[Y]$ exist!) ∎

$$\text{cov}(X,X) = E[X^2] - E[X]E[X] = \text{var}(X)$$

$$\text{cov}(X,aY) = E[XaY] - E[X]E[aY] = aE[XY] - aE[X]E[Y] = a(E[XY] - E[X]E[Y]) = a\text{cov}(X,Y)$$

$$\text{cov}(X,Y) = E[XY] - E[X]E[Y] = E[YX] - E[Y]E[X] = \text{cov}(Y,X)$$

Exercise 2. Prove for any X and Y independent or not, $\text{var}(X+Y) = \text{var}(X) + 2\text{cov}(X,Y) + \text{var}(Y)$. ∎

From lecture 1:

$$\text{var}[X+Y] = E[X^2] + 2E[XY] + E[Y^2] - (EX)^2 - 2E[X]E[Y] - (EY)^2$$

$$= E[X^2] + 2E[XY] + E[Y^2] - (EX)^2 - 2E[X]E[Y] - (EY)^2$$

$$\text{var}(X+Y) = \text{var}(X) + 2E[XY] - 2E[X]E[Y] + \text{var}(Y)$$

$$\text{var}(X+Y) = \text{var}(X) + 2\text{cov}[X,Y] + \text{var}(Y)$$

Exercise 3. Pick $a = 1/3$, and try adding $(1/3)^0 + (1/3)^1 + (1/3)^2 + \cdots = 1 + 1/3 + 1/9 + 1/27 + \cdots$. In other words, add successive terms and note the behavior. What does the formula say the limit ought to be? ∎

Use R:

```
sum <- 0
sums <- c()
for(ii in 0:15){
        sum <- sum + (1/3)^ii
        sums <- c(sums, sum)
}
```

Number of Terms	Sum
1	1.0000
2	1.3333
3	1.4444
4	1.4815
5	1.4938
6	1.4979
7	1.4993
8	1.4998
9	1.4999
10	1.5000
11	1.5000
12	1.5000
13	1.5000
14	1.5000
15	1.5000
16	1.5000

The limit ought to be $1/(1-1/3) = 1.5$

Exercise 4. Use the geometric series formula to show that the sum of the geometric probabilities is 1. Pay attention to the fact that the geometric probabilities start at 1, not 0. ∎

$$\sum_y P(Y=y) = \sum_{y=1}^{\infty} p(1-p)^{y-1} = p\sum_{u=0}^{\infty}(1-p)^u = p\frac{1}{1-(1-p)} = 1$$

Exercise 5. Suppose that each time a case of disease is introduced into a region, $a < 1$ secondary cases are produced from each case on average. Show that the size of the n generation is a^n. Don't worry about what fractions of a person mean; interpret these in some average sense. Show that the total size of the outbreak—adding all cases together including the index case—is $1/(1-a)$. ∎

Proof by induction:

Step 1: Show the base case is true.

In the first generation the expected number of cases is a (given in the problem).

Step 2: If the expected size of the nth generation is a^n, show the expected size of the $n+1$th generation is a^{n+1}.

Therefore, we have proved that the size of the nth generation is a^n.

The total outbreak size is given by:

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}$$

Exercise 6. Suppose that on average a case of disease causes $R > 1$ secondary cases when no one is vaccinated or immune. Suppose now that a fraction f is vaccinated with complete protection. Discuss the epidemiological credibility of the assumption that now each case would cause (on average) $R(1-f)$ secondary cases (i.e. when the coverage fraction is f). (Do you see the expectation of a binomial possibly concealed in here?) Assume that with coverage f, each case causes $R(1-f)$ secondary cases, and that $R(1-f) < 1$; compute the total expected size of an outbreak, counting all generations. What is the critical coverage level needed for $R(1-f) < 1$? ∎

We are making the assumption that each person's vaccination status is independent of another's. This may not be a valid assumption in the real world, where non-vaccinated individuals may cluster at certain schools or in certain neighborhoods.

$$R(1-f) < 1$$

$$1 - f < \frac{1}{R}$$

$$f > 1 - \frac{1}{R}$$

The critical coverage level is $1 - \frac{1}{R}$.

Exercise 7. Use Newton's formula

$$(a+b)^N = \sum_{i=0}^{N} \binom{N}{i} a^i b^{N-i}$$

to prove that the probability generating function for a binomial is $(up + (1-p))^N$. ∎

Let g(u) be the PGF for a binomial.

$$g(u) = E[u^X] = \sum_{x=0}^{N} u^x \binom{N}{x} p^x (1-p)^{N-x}$$

$$= \sum_{x=0}^{N} \binom{N}{x} (up)^x (1-p)^{N-x}$$

$$= (up + (1-p))^{N-x}$$

Exercise 8. Suppose a needle is discarded with probability 0.2 after every usage; each disposal decision is independent of the past. What is the expected number of usages? ∎

$1/0.2 = 5$

Exercise 9. Let Y be geometric with parameter p. Prove that the conditional distribution of $Y - Y_0$ given $Y > Y_0$ for any Y_0 is geometric with parameter p also. ∎

We can do this by showing: $E[u^Y] = E[u^{Y-Y_0}|Y > Y_0]$.

$$E[u^Y] = \sum_{i=0}^{\infty}(1-p)^{i-1}pu^i = \frac{up}{1-(1-p)u}$$

$$E[u^{Y-Y_0}|Y > Y_0] = \sum_{i=Y_0+1}^{\infty}(1-p)^{i-Y_0-1}pu^{i-Y_0}$$

$$= \sum_{j=1}^{\infty}(1-p)^{j-1}pu^j$$

$$= \frac{up}{1-(1-p)u}$$

Since both distributions have the same generating function, they are the same distribution.

Exercise 10. Intuitively we know the mean of the Poisson should be λ, from the derivation. But what if something untoward happened when we took the limit? Show that if Y is Poisson with parameter λ, then $E[Y] = \lambda$. ∎

$$E[Y] = \sum_{i=0}^{\infty} i\frac{e^{-\lambda}\lambda^i}{i!} = \lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{i\lambda^{i-1}}{i!}$$

$$= \lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{d}{d\lambda}\frac{\lambda^i}{i!}$$

$$= \lambda e^{-\lambda}\frac{d}{d\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!}$$

$$= \lambda e^{-\lambda}\frac{d}{d\lambda}e^{\lambda} = \lambda$$

Exercise 11. Suppose that accidents occur according to a Poisson distribution with mean 1. What is the probability of zero accidents? Suppose the mean is 0.1 or 10; what is the probability of zero accidents? ∎

Use R:

```r
dpois(0,1)

## [1] 0.3678794
```

```
dpois(0,.1)

## [1] 0.9048374

dpois(0,10)

## [1] 4.539993e-05
```

Exercise 12. Suppose that HIV-negative participants in a trial acquire new partners at random according to a Poisson distribution with mean λ. Supposing that the prevalence of infection is p and the probability of infection given an infected partner is β, what is the distribution of the number of *sufficient exposures*? What is the probability of zero sufficient exposures? What is the probability of the person becoming infected (i.e. of having one or more sufficient exposures)? ∎

$$P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!} = \frac{e^{-\beta p}(\beta p)^i}{i!}$$

Probability of 0 sufficient exposures:

$$P(X = 0) = \exp(-\beta p)$$

Probability of at least 1 sufficient exposure:

$$1 - P(X = 0) = 1 - \exp(-\beta p)$$

Computer Exercise 13. Please use the above R code to simulate the IAR(1) process with (a) $\lambda = 10$ and $p = 0$ (no survival), $p = 0.3$, and $p = 0.8$. Part (b): Also, use $\lambda = 200$, with $p = 0.1$ in one run and $p = 0.99$ in another; please comment on the differences. Please try, in an exploratory way, different values of λ and p; feel free to set `xz <- 1:500` to go out to time 500 or even longer. Plot your results; comment on the appearance of the graph (smoothness, initial rise). Only turn in 3–4 graphs. ∎

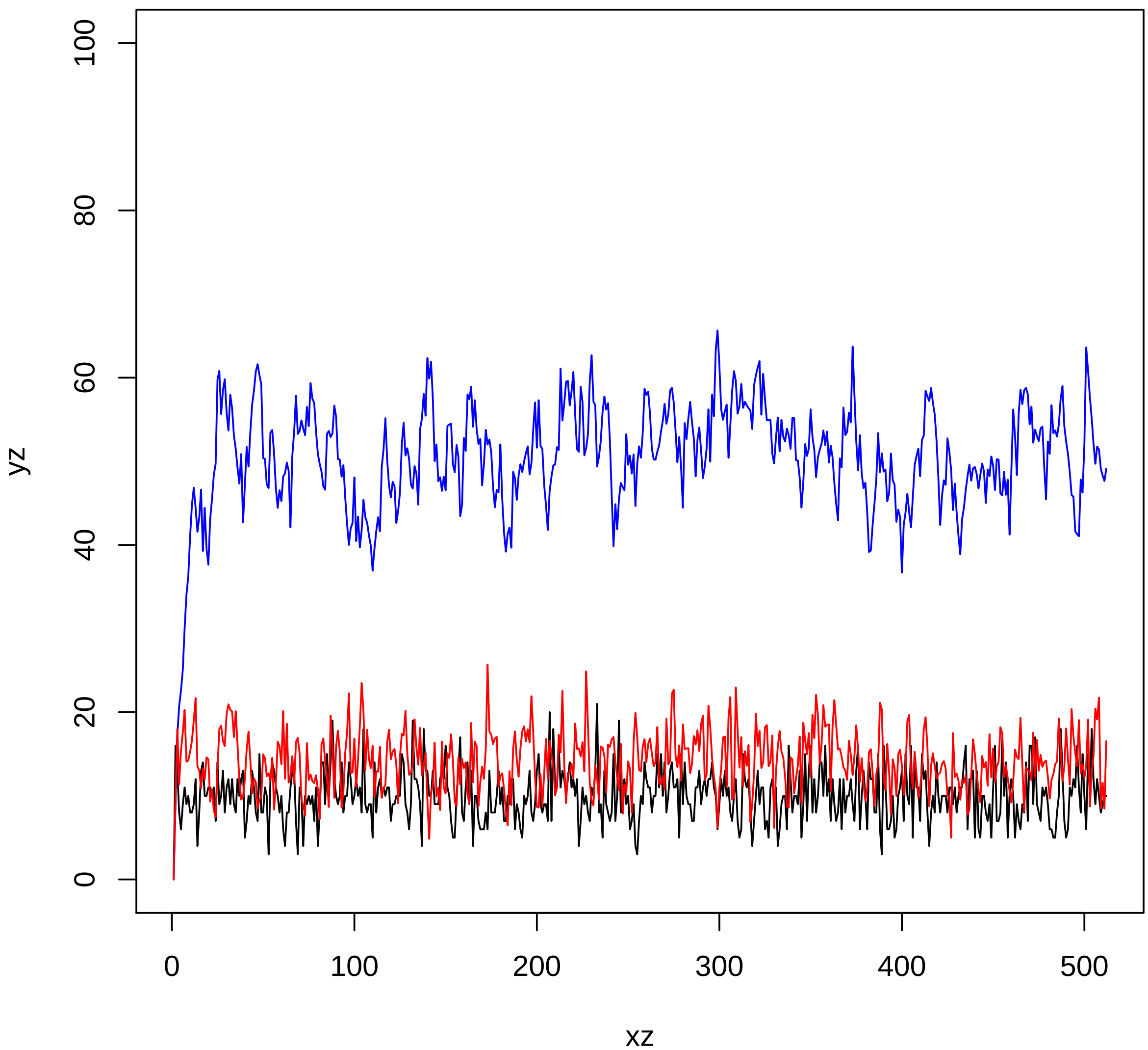
Load in given code:

```
iar1sim <- function(len,lambda,pp,starting=0) {
  if (len <= 0 || len > 65536) {
    stop("length bad")
  }
  if (pp<0 || pp>1.0) {
    stop("p bad")
  }
```

```r
  if (lambda < 0) {
    stop("lambda bad")
  }
  answer <- rep(NA,len)
  answer[1] <- starting
  cur <- starting
  for (ii in 2:len) {
    cur <- pp*cur + rpois(1,lambda=lambda)
    answer[ii] <- cur
  }
  answer
}
```

a)

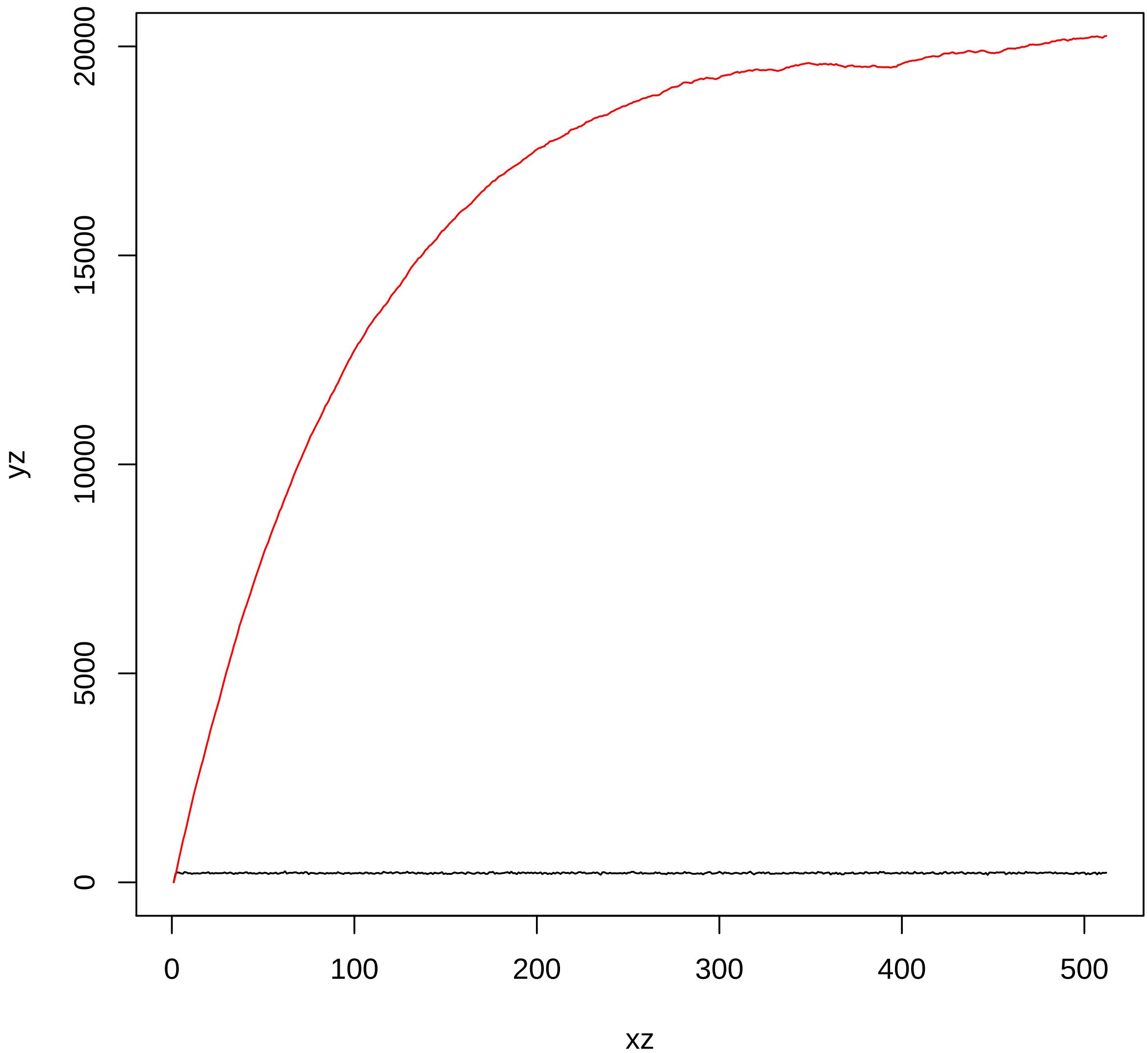
```r
xz <- 1:512
yz <- iar1sim(length(xz), 10, 0)
yz2 <- iar1sim(length(xz), 10, 0.3)
yz3 <- iar1sim(length(xz), 10, 0.8)
plot(xz,yz,type="l",ylim=c(0,100))
points(xz,yz3,type="l",col="blue")
points(xz,yz2,type="l",col="red")
```



b)

```
xz <- 1:512
yz <- iar1sim(length(xz), 200, 0.1)
yz2 <- iar1sim(length(xz), 200, 0.99)
plot(xz,yz,type="l",ylim=c(0,2e4))
```

```r
points(xz,yz2,type="l",col="red")
```



Exercise 14. How long did this assignment take you? ∎