Solutions to Lab 4

©2005 Ben Bolker, modified at some places by Alejandro Morales & Bob Douma 2017

November 14, 2018

Exercise 1: For the binomial distribution with 10 trials and a success probability of 0.2:

- Pick 8 random values and sort them into increasing order (if you `set.seed(1001)` beforehand, you should get $X = 0$ (twice), $X = 2$ (4 times), and $X = 4$ and $X = 5$ (once each)).
- Calculate the probabilities of getting 3, 4, or 5 successes first by hand (See Ch.4) and check it with the computer. Answer:

The equation for the binomial distribution is $N!/(x!(N-x)!)p^x(1-p)^x$ Thus we get $10!/(3!(10-3)!)0.2^3(1-0.2)^(10-3)$ $10!/(4!(10-4)!)0.2^4(1-0.2)^(10-4)$ $10!/(5!(10-5)!)0.2^5(1-0.2)^(10-5)$

```
factorial(10)/(factorial(3) * factorial((10 - 3))) * 0.2^3 *
    (1 - 0.2)^(10 - 3)
```

```
## [1] 0.2013266
```

```
factorial(10)/(factorial(4) * factorial((10 - 4))) * 0.2^4 *
    (1 - 0.2)^(10 - 4)
```

```
## [1] 0.08808038
```

```
factorial(10)/(factorial(5) * factorial((10 - 5))) * 0.2^5 *
    (1 - 0.2)^(10 - 5)
```

```
## [1] 0.02642412
```

```
dbinom(3:5, size = 10, prob = 0.2)
```

```
## [1] 0.20132659 0.08808038 0.02642412
```

- Calculate the probability of getting 5 or more successes. Answer:

```
pbinom(4, size = 10, prob = 0.2, lower.tail = FALSE)
```

```
## [1] 0.0327935
```

- What tail values would you use to test against the (two-sided) null hypothesis that `prob=` 0.2 at the 95% level? (Use `qbinom()`

to get the answer, and use `pbinom(0:10,size=10,prob=0.2)` and `pbinom(0:10,size=10,prob=0.2,lower.tail=FALSE)` to check that your answer makes sense.

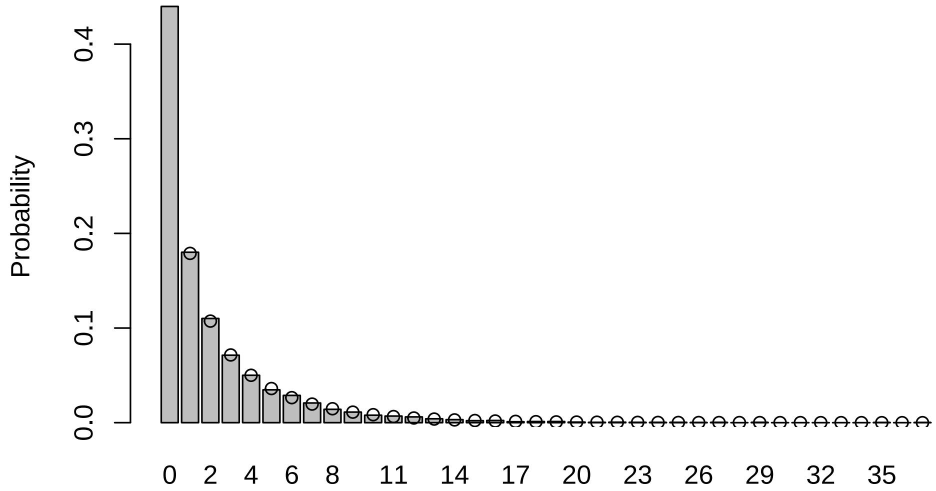
```
qbinom(c(0.025, 0.975), prob = 0.2, size = 10)
```

```
## [1] 0 5
```

The actual answer based on these results (0,5) is that we will not be able to detect a deviation below 0.2 with only 10 samples; 6 or more successes would suggest a significantly greater probability. (The probability of getting 5 or more successes, or `pbinom(4,size=10,prob=0.2, lower.tail=FALSE)` is 0.032, which does not attain the 2.5% level we are looking for in the upper tail. The probability of 6 or more successes, `pbinom(5,size=10,prob=0.2,lower.tail=FALSE)`, is 0.006. We would need a sample size of 17 to be able to detect a probability significantly below 0.2.)

Exercise 2*: Pick 10,000 negative binomial deviates with $\mu = 2$, $k = 0.5$ (using `rnbinom()`). Pick one of the ways above to draw the distribution. Check that the mean and variance agree reasonably well with the theoretical values. Add points representing the theoretical distribution to the plot.

```
mu = 2
k = 0.5
x = rnbinom(10000, mu = mu, size = k)
tx = table(factor(x, levels = 0:max(x)))/10000
b1 = barplot(tx, ylab = "Probability")
points(b1, dnbinom(0:max(x), mu = mu, size = k), pch = 1)
```



```r
mean(x)
```

```
## [1] 2.0466
```

```r
var(x)
```

```
## [1] 10.33386
```

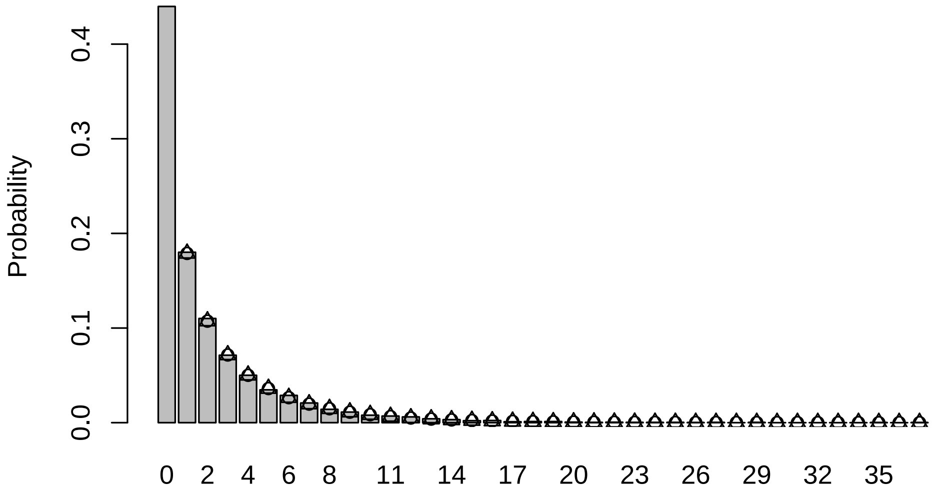
```r
mu
```

```
## [1] 2
```

```r
mu * (1 + mu/k)
```

```
## [1] 10
```

```r
# the
# alternative
# parameterisation
p = 1/(1 + mu/k)
n = k

b1 = barplot(tx, ylab = "Probability")
points(b1, dnbinom(0:max(x), mu = mu, size = k), pch = 1)
points(b1, dnbinom(0:max(x), prob = p, size = k), pch = 2)
```



2. Averaging distributions

Exercise 3: figure out what it means that mean(tabdat) equals r mean(tabdat).

```
dat = c(5, 6, 5, 7, 5, 8)
dat
```

```
## [1] 5 6 5 7 5 8
```

```
tabdat = table(dat)
tabdat
```

```
## dat
## 5 6 7 8
## 3 1 1 1
```

```
mean(tabdat)
```

```
## [1] 1.5
```

2.1. Jensen's inequality

Exercise 4: In statistical models, you often estimate the mean effect of a given treatment. Find out what the effect of Jensen's inequality is on a series of log-tranformed datapoints with respect to the estimated mean.

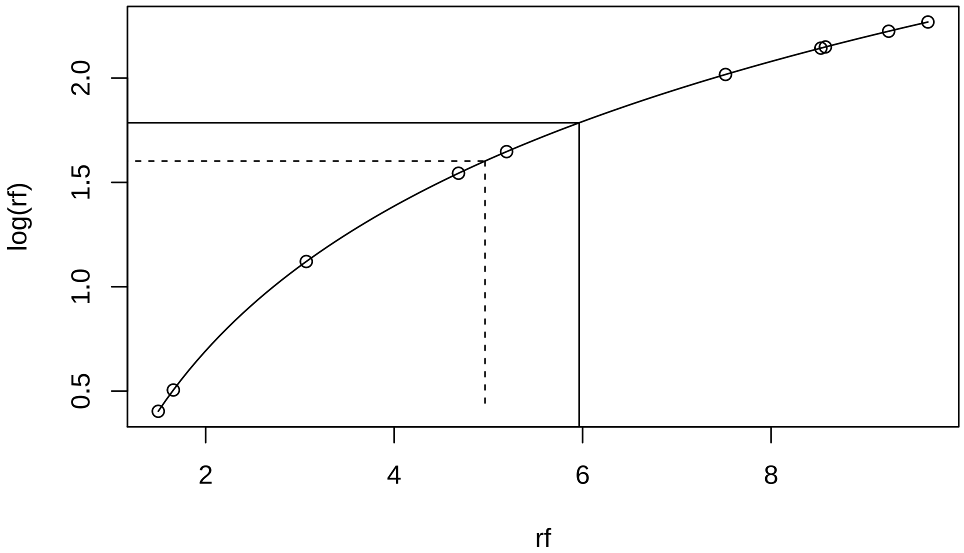
```
rf = runif(10, min = 0, max = 10)
mean(rf)
```

```
## [1] 5.963339
```

```
exp(mean(log(rf)))
```

```
## [1] 4.964574
```

```
plot(log(rf) ~ rf)
curve(log(x), add = T)
segments(x0 = 0, y0 = log(mean(rf)), x1 = mean(rf), y1 = log(mean(rf)),
    lty = 1)
segments(x0 = mean(rf), y0 = 0, x1 = mean(rf), y1 = log(mean(rf)),
    lty = 1)

segments(x0 = 0, y0 = mean(log(rf)), x1 = exp(mean(log(rf))),
    y1 = mean(log(rf)), lty = 2)
segments(x0 = exp(mean(log(rf))), y0 = mean(log(rf)), x1 = exp(mean(log(rf))),
    y1 = min(log(rf)), lty = 2)
```



This shows that the log of the mean of x is not the same as the mean of the log of x

Exercise 5* :

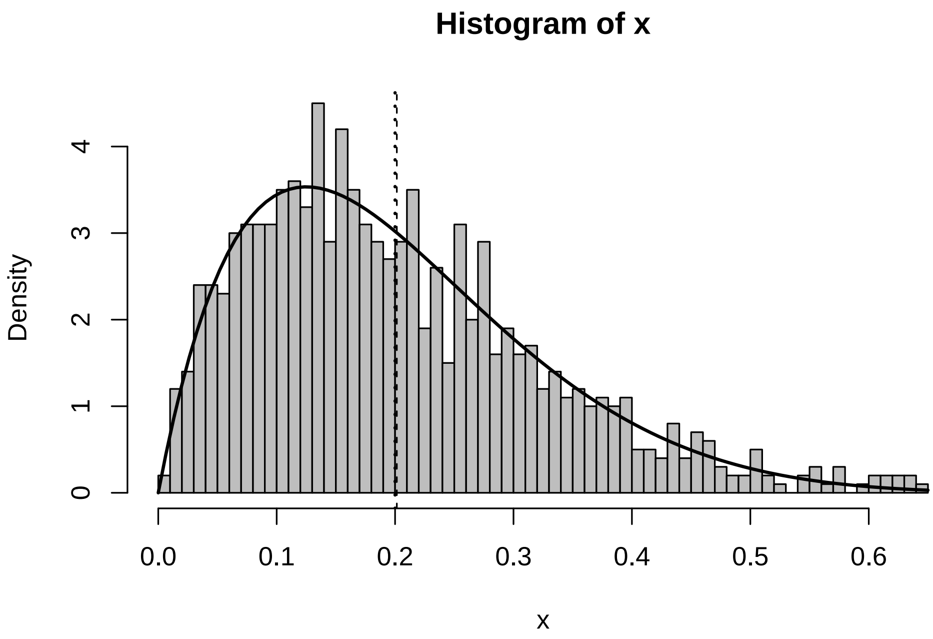
Morris (1997) gives a definition of the beta function that is different from the standard statistical parameterization.

Find expressions for P and θ in terms of a and b

Based just on the expressions in the normalization constant $\Gamma(a+b)/(\Gamma(a)\Gamma(b))$ for the standard parameterization, $\Gamma(\theta)/(\Gamma(\theta P)\Gamma(\theta(1-P))))$ gives $\theta = a+b$, $P = a/(a+b)$ or conversely $a = \theta P$, $b = \theta(1-P)$. In this parameterization, P is the mean proportion/ number of successes/etc. and θ governs the width of the distribution

```
my_rbeta = function(n, theta, P) {
    rbeta(n, shape1 = theta * P, shape2 = theta * (1 - P))
}

my_dbeta = function(x, theta, P) {
    dbeta(x, shape1 = theta * P, shape2 = theta * (1 - P))
}
```

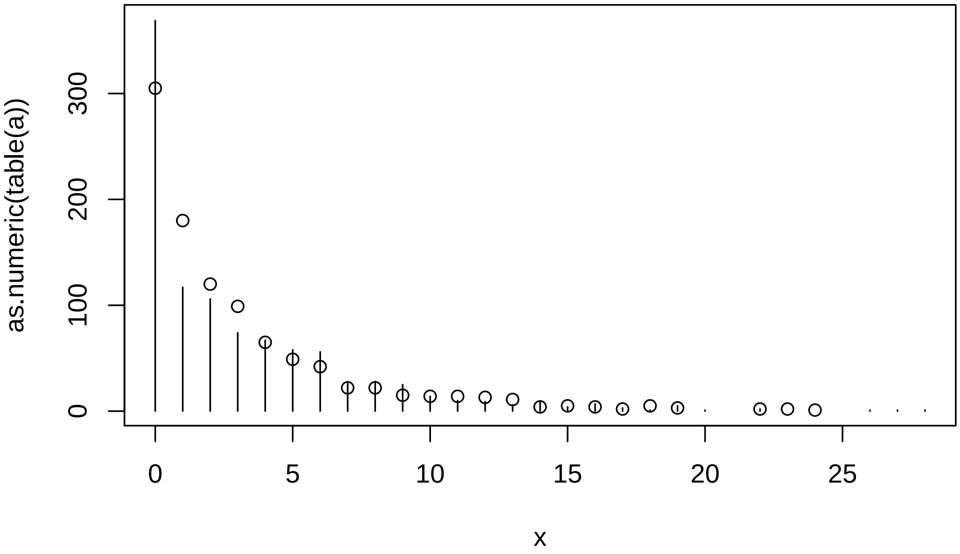
```r
x = my_rbeta(1000, theta = 10, P = 0.2)
hist(x, breaks = 50, prob = TRUE, col = "gray")
curve(my_dbeta(x, theta = 10, P = 0.2), add = TRUE, lwd = 2)
abline(v = 0.2, lwd = 2, lty = 3)
abline(v = mean(x), lty = 2)
```



Exercise 6: Check graphically that these functions actually work. For instance, you could compare the results with a negative binomial function with the same mean and variance as the data.

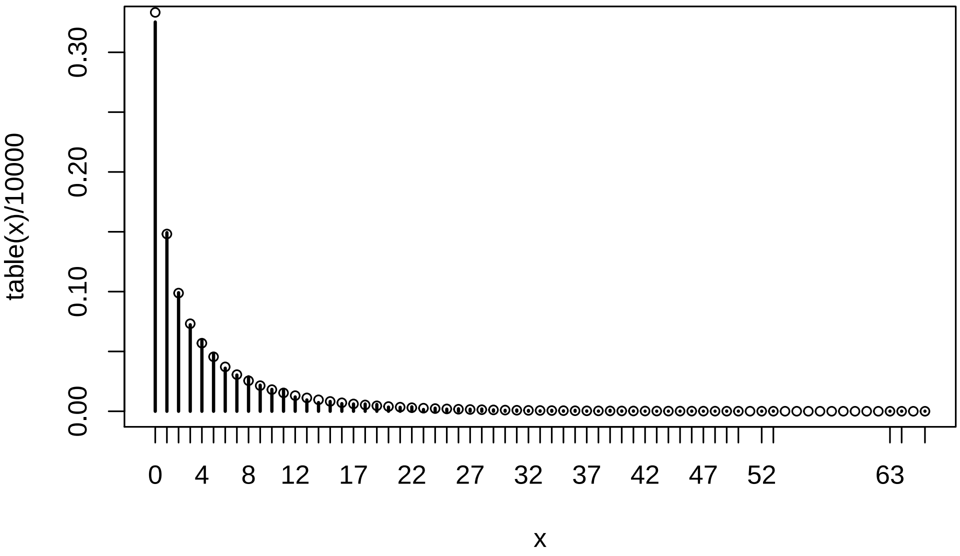
```r
rzinbinom = function(n, mu, size, zprob) {
    ifelse(runif(n) < zprob, 0, rnbinom(n, mu = mu, size = size))
}

a = rzinbinom(1000, mu = 4, size = 1, zprob = 0.2)

mean.a = mean(a)
var.a = var(a)
size = 1/(((var.a - mean.a))/mean.a^2)

a1 = rnbinom(1000, mu = mean.a, size = size)
```

```
x = as.numeric(names(table(a)))
plot(as.numeric(table(a)) ~ x, type = "h")
x = as.numeric(names(table(a1)))
points(as.numeric(table(a1)) ~ x, type = "p")
```



Exercise 7*: generate 10,000 values from a gamma-Poisson compounded distribution with parameters shape=$k = 0.5$, scale=$\mu/k = 4/0.5 = 8$ and demonstrate that it's equivalent to a negative binomial with the appropriate μ and k parameters.

```
mu = 4
k = 0.5
x = rpois(10000, rgamma(10000, shape = k, scale = mu/k))
plot(table(x)/10000)
points(0:max(x), dnbinom(0:max(x), mu = mu, size = k), cex = 0.75)
```



6. Choosing probability distributions

Exercise 8 Reload the six datasets and choose for each dataset one or two candidate probability distributions. *hint*: ask yourself a number of questions: Do I have counts (integer) or continuous (real) values? Do I have positive values only or also negative values?

dataset 1 A light response curve typically present photosynthesis values. This could be net or gross photosynthesis. In case of the former, it can include negative values. In any case the values are continuous. I would opt for a normal or gamma distribution

dataset 2 Intake rate of a predator typically has units “number of prey” or “prey denisty” per unit time. It can only be positive and either discrete numbers or continous numbers. The numbers in the file are continuous so I would model this througha a gamma or lognormal

dataset 3 No information is given what type of allometric relationship is in the file. As the y contains integer values, I would go for a poisson or negative binomial.

dataset 4 This dataset contains of measurements of population size over time. The data shows integer values, so presumably these are numbers and I would opt for a poisson or negative binomial.

dataset 5 No information is given what kind of data this is. The y columns shows integer values only. Values below 0 are lacking, so I would opt for a poisson or negative binomial.

dataset 6 Presence/absence 0,1 data can be modelled with a binomial or beta-binomial distribution.