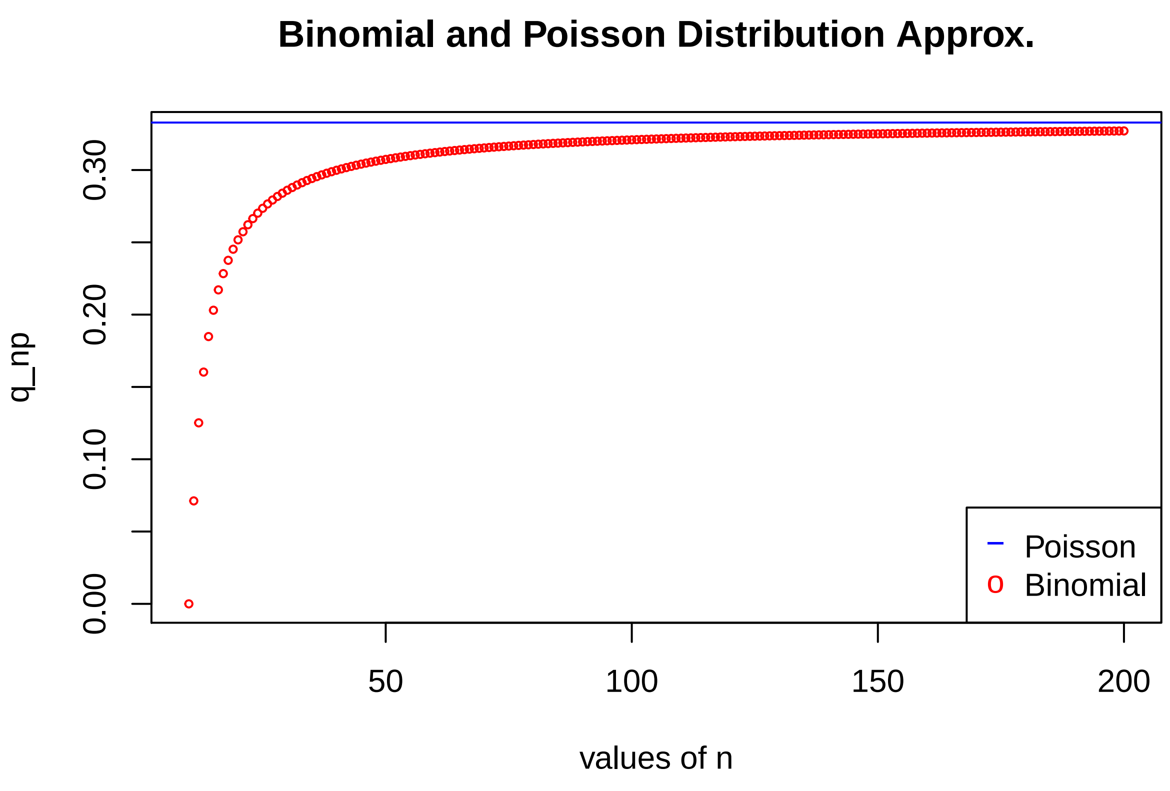
DSC 462 Assignment 3

Shouman Das

October 26, 2017

Q1

```r
lambda = 10
n = 10:200
p = lambda/n
q_np = pbinom(8, n, p)
plot(n,q_np, main = "Binomial and Poisson Distribution Approx.",
     xlab="values of n", ylab="q_np", cex=.5, col = 'red')
abline(h=ppois(8,10), col="blue")
legend("bottomright",c("Poisson", "Binomial"), col = c("blue","red"), pch = c('-o'))
```




```r
pois = ppois(8,10)
n[which(abs(pois-q_np)<0.01)[1]]
```

```
## [1] 119
```

So the smallest n for which $|q_{n,p} - q_\lambda| < 0.01$ is 119.

Q2

(a) $X_1, X_2 \sim geom(p)$. So for $s \geq 2$ we have,

$$\begin{aligned}
P(X_1 + X_2 = s) &= \sum_{i=1}^{s-1} P(X_1 = i, X_2 = s - i) \\
&= \sum_{i=1}^{s-1} P(X_1 = i)P(X_2 = s - i) \\
&= \sum_{i=1}^{s-1} (1-p)^{i-1} p(1-p)^{s-i-1} p \\
&= (s-1)p^2(1-p)^{s-2}
\end{aligned}$$

Also, we have $P(X_1 = x, X_1 + X_2 = s) = P(X_1 = x, X_2 = s - x) = P(X_1 = x)P(X_2 = s - x) = p(1-p)^{x-1}p(1-p)^{s-x-1} = p^2(1-p)^{s-2}$. So,

$$P(X_1 = x | X_1 + X_2 = s) = \frac{P(X_1 = x, X_1 + X_2 = s)}{P(X_1 + X_2 = s)} = \frac{1}{s-1}$$

(b) Whenever $1 \leq x \leq s-1$, we see that $p_X(x)$ doesn't depend on x or p. In other case it is zero.

Q3

By theorem 5.3,

$$Odds(A|X = x) = \frac{P(X = x|A)}{P(X = x|A^c)} \times Odds(A)$$

which is equivalent to $Odds(A|X = x) \times P(X = x|A^c) = P(X = x|A) \times Odds(A)$. Also,

$$P(X = x|A) = \binom{4}{x} 0.5^x 0.5^{4-x} = \binom{4}{x} 0.5^4$$

and

$$P(X = x|A^c) = \binom{2}{x} 0.9^x 0.1^{2-x}$$

For $x = 0$, $Odds(A|X = 0) \times 0.1^2 = 0.5^4 \times Odds(A) \iff Odds(A|X = 0) \times 0.16 = Odds(A)$.

For $x = 1$, $Odds(A|X = 1) \times (2)(0.9)(0.1) = (4)0.5^4 \times Odds(A) \iff Odds(A|X = 1) \times 0.72 = Odds(A)$.

For $x = 2$, $Odds(A|X = 1) \times (0.9)^2 = (6)0.5^4 \times Odds(A) \iff Odds(A|X = 1) \times 2.16 = Odds(A)$.

For $x = 3$, $Odds(A|X = 1)(0) = (4)0.5^4 \times Odds(A)$.

For $x = 4$, $Odds(A|X = 1)(0) = (1)0.5^4 \times Odds(A)$.

Now we know that $Odds(A^c) = 1/Odds(A)$ and $Odds(A^c|X = x) = 1/Odds(A|X = x)$. So for $x = 2$, we see that the evidence of the form $\{X = x\}$ increase the odds that A does **not** occur.

Q4

		True infection		
		Positive	Negative	Total
Test	Positive	256	12	268
	Negative	29	208	237
	Total	285	220	505

1. $Sensitivity = \frac{256}{256+29} = 0.8982456$.

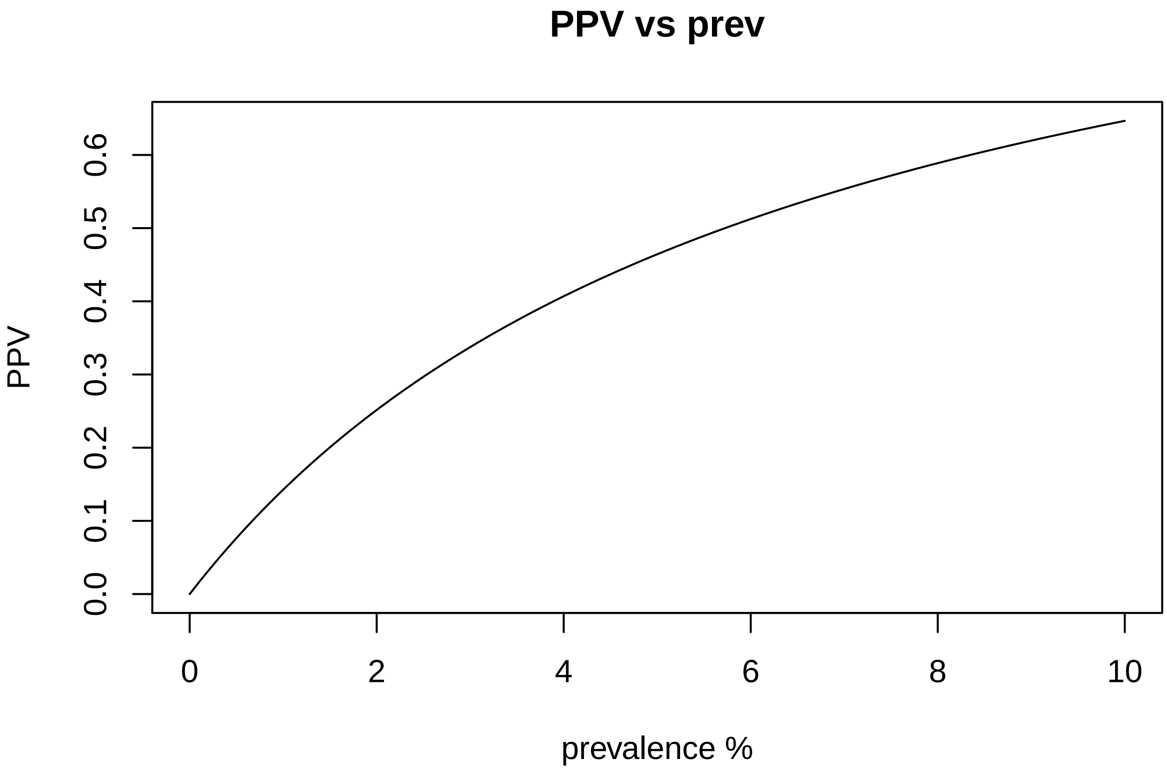
```
sens = 256/285
```

And $specificity = \frac{208}{208+12} = 0.9454545$.

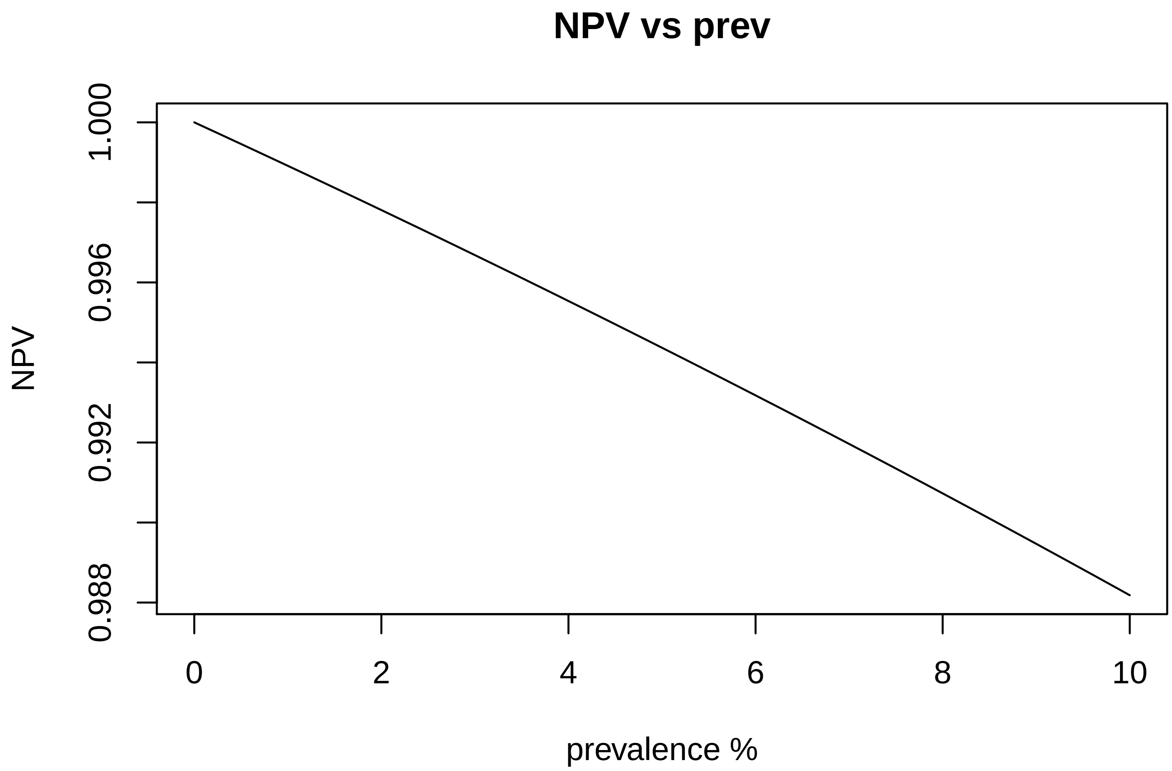
```
spec = 208/220
```

2.

```
prev = seq(0,10,0.1)/100
PPV = sens*prev/(sens*prev + (1-spec)*(1-prev))
NPV = spec*(1-prev)/(spec*(1-prev) + (1-sens)*prev)
plot(prev*100, PPV,type = 'l',xlab = 'prevalence %', main = "PPV vs prev")
```




```
plot(prev*100,NPV,type = 'l',xlab = 'prevalence %', main = "NPV vs prev")
```



3. Directly from data,

$$prev = \frac{285}{505} = 0.5643564$$

$$PPV = \frac{256}{268} = 0.9552239$$

$$NPV = \frac{208}{237} = 0.8776371$$

```r
285/505
```

```
## [1] 0.5643564
```

```r
256/268
```

```
## [1] 0.9552239
```

```r
208/237
```

```
## [1] 0.8776371
```

We can see that from the direct computation from data, our PPV is much higher and NPV is much lower.

Q5

(a) First note that, $\frac{d^k e^{tx}}{dt^k}|_{t=0} = x^k e^{tx}|_{t=0} = x^k$. Now assuming X is continuous random variable (discontinuous case is similar), we have

$$\frac{d^k M_X(t)}{dt^k}\Big|_{t=0} = \frac{d^k E[e^{tX}]}{dt^k}\Big|_{t=0} = \frac{d^k}{dt^k}\left[\int_{-\infty}^{\infty} e^{tx} f(x)dx\right]\Big|_{t=0}$$

Since we can interchange the order of integration and derivative, we get

$$\frac{d^k M_X(t)}{dt^k}\Big|_{t=0} = \int_{-\infty}^{\infty} \frac{d^k e^{tx}}{dt^k}\Big|_{t=0} f(x)\ dx = \int_{-\infty}^{\infty} x^k f(x)\ dx = E[X^k]$$

(b) $M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}]$. Since X, Y are independent, e^{tX}, e^{tY} are also independent. So, $E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}]$. Therefore, we have,

$$M_{X+Y}(t) = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t).$$

(c) We have pmf, $p_X(i) = \binom{n}{i}p^i(1-p)^{n-i}$. So

$$E[e^{tX}] = \sum_{i=0}^{n}\binom{n}{i}e^{ti}p^i(1-p)^{n-i}] = (1-p+pe^t)^n.$$

(d) If $p = q$ and $X \sim bin(n,p), Y \sim bin(m,q)$, then $M_{X+Y}(t) = M_X(t)M_Y(t) = (1-p+pe^t)^n(1-q+qe^t)^m = (1-p+pe^t)^{m+n}$. Since all the moments are finite, we can say $X+Y \sim bin(n+m,p)$.

Similarly, if $X+Y$ is a binomial random variable, then by (c), $M_{X+Y}(t) = (1-r+re^t)^N$ for some nonnegative integer N and $r \in [0,1]$. But by (b), $M_{X+Y}(t) = M_X(t)M_Y(t) = (1-p+pe^t)^n(1-q+qe^t)^m$. Since this has to be true for all t, we can get $p = q$ and $N = n+m$.