APPROXIMATING THE POISSON DISTRIBUTION USING THE CENTRAL LIMIT THEOREM

Suppose that $Y_1, Y_2, \ldots, Y_n$, are independent random variables that have the same distribution, and have expectation μ and variance σ^2. The Central Limit Theorem states that the random variable

$$U_n = \frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma}$$

converges in distribution to a *standard Normal distribution*. We could also write

$$U_n = \frac{(\overline{Y}_n - \mu)}{\sigma/\sqrt{n}} \qquad \text{or} \qquad U_n = \frac{\left(\sum_{i=1}^{n} Y_i - n\mu\right)}{\sqrt{n}\sigma}.$$

Thus, when n is large, we can **approximate** the distribution of $\overline{Y}_n$, or of

$$S_n = \sum_{i=1}^{n} Y_i$$

using a Normal distribution, irrespective of the actual distribution of $Y_1, \ldots, Y_n$. This Theorem, along with properties of the Poisson distribution, allow the approximation of the Poisson pmf and cdf when λ is large. Specifically, if $Y \sim Poisson(\lambda)$, then $\mathbb{E}[Y] = \mathbb{V}[Y] = \lambda$, and we consider the transformation

$$X = \frac{Y - \lambda}{\sqrt{\lambda}}.$$

As $\lambda \longrightarrow \infty$, X converges in distribution to a standard Normal random variable, that is

$$\lim_{\lambda \longrightarrow \infty} F_X(x) = \lim_{\lambda \longrightarrow \infty} F_Y\left(\sqrt{\lambda}x + \lambda\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du$$

Also if

$$X = \frac{Y - \lambda}{\sqrt{\lambda}} = \frac{Y}{\sqrt{\lambda}} - \sqrt{\lambda}.$$

then for the mgf, we have directly that

$$m_X(t) = e^{-\sqrt{\lambda}t} m_Y(t/\sqrt{\lambda}) = \exp\left\{-\sqrt{\lambda}t + \lambda(e^{t/\sqrt{\lambda}} - 1)\right\}.$$

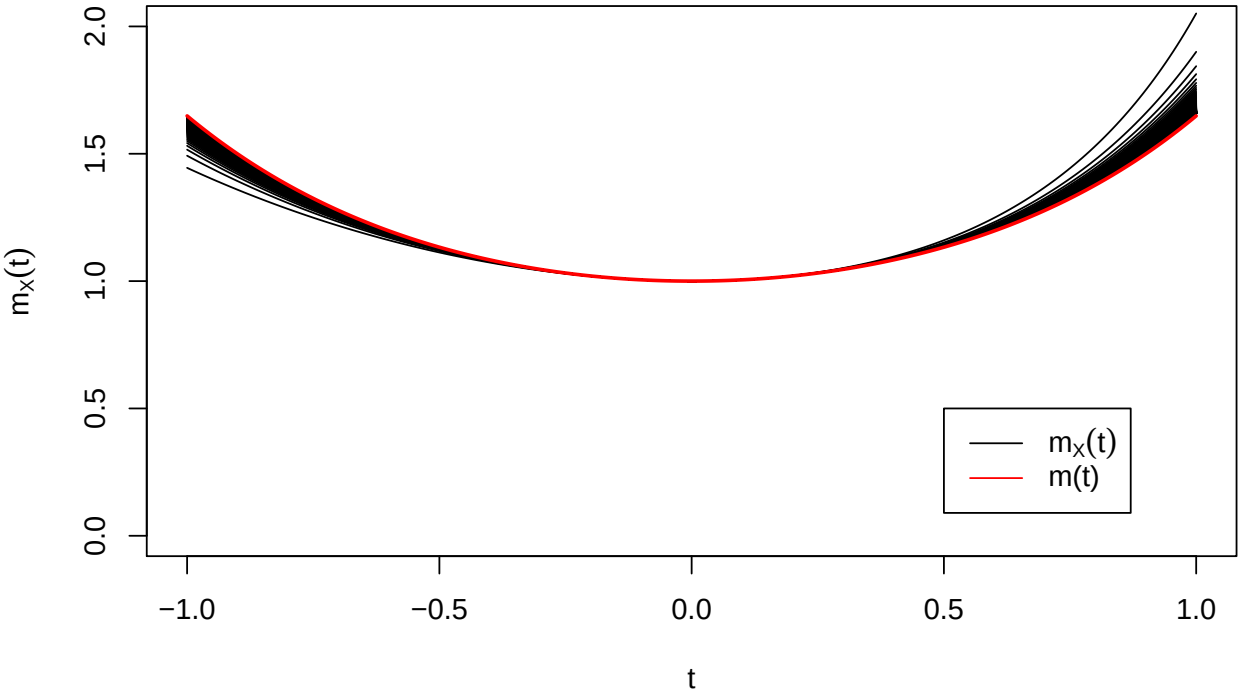
Using the series expansion for the exponent term, we have that

$$m_X(t) = \exp\left\{-\sqrt{\lambda}t + \lambda\left(\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{6\lambda^{3/2}} + \cdots\right)\right\} = \exp\left\{\frac{t^2}{2} + \frac{t^3}{6\lambda^{1/2}} + \cdots\right\} \longrightarrow \exp\left\{\frac{t^2}{2}\right\}$$

as $\lambda \longrightarrow \infty$. This final form is the mgf of the standard Normal distribution. The convergence of the mgf is demonstrated in the plot below: the black lines plot the function $m_X(t)$ for $-1 \leq t \leq 1$, and the red line shows the limiting function

$$m(t) = \exp\left\{\frac{t^2}{2}\right\}.$$

```
set.seed(1010)
tv<-seq(-1,1,by=0.001)
lam<-seq(1,1000,by=1)
slam<-sqrt(lam)
yv<-exp(-slam[1]*tv)*exp(lam[1]*(exp(tv/slam[1])-1))
plot(tv,yv,type='l',ylim=range(0,2),xlab='t',ylab=expression(m[X](t)))
for(i in 2:1000){
        yv<-exp(-slam[i]*tv)*exp(lam[i]*(exp(tv/slam[i])-1))
        lines(tv,yv)
}
lines(tv,exp(tv^2/2),col='red',lwd = 2)
legend(0.5,0.5,c(expression(m[X](t)),'m(t)'),col=c('black','red'),lty=1)
```



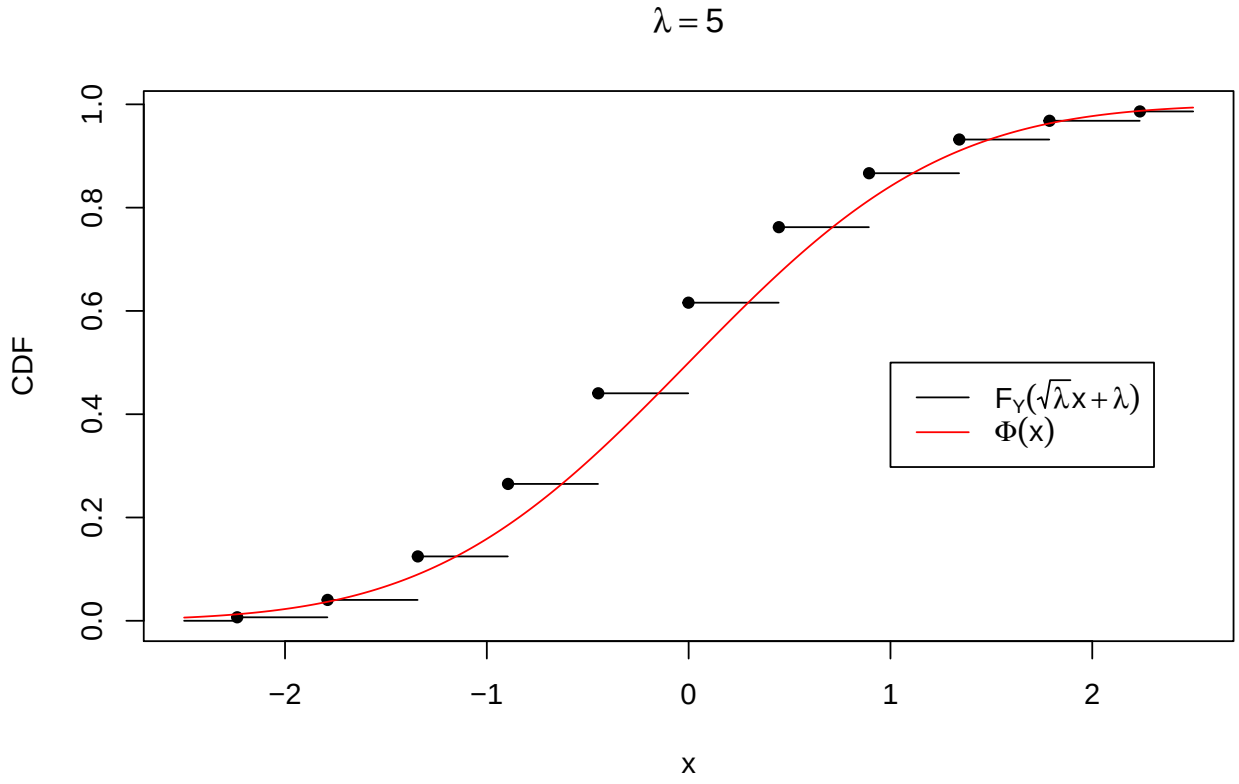
We can assess the accuracy of the cdf approximation

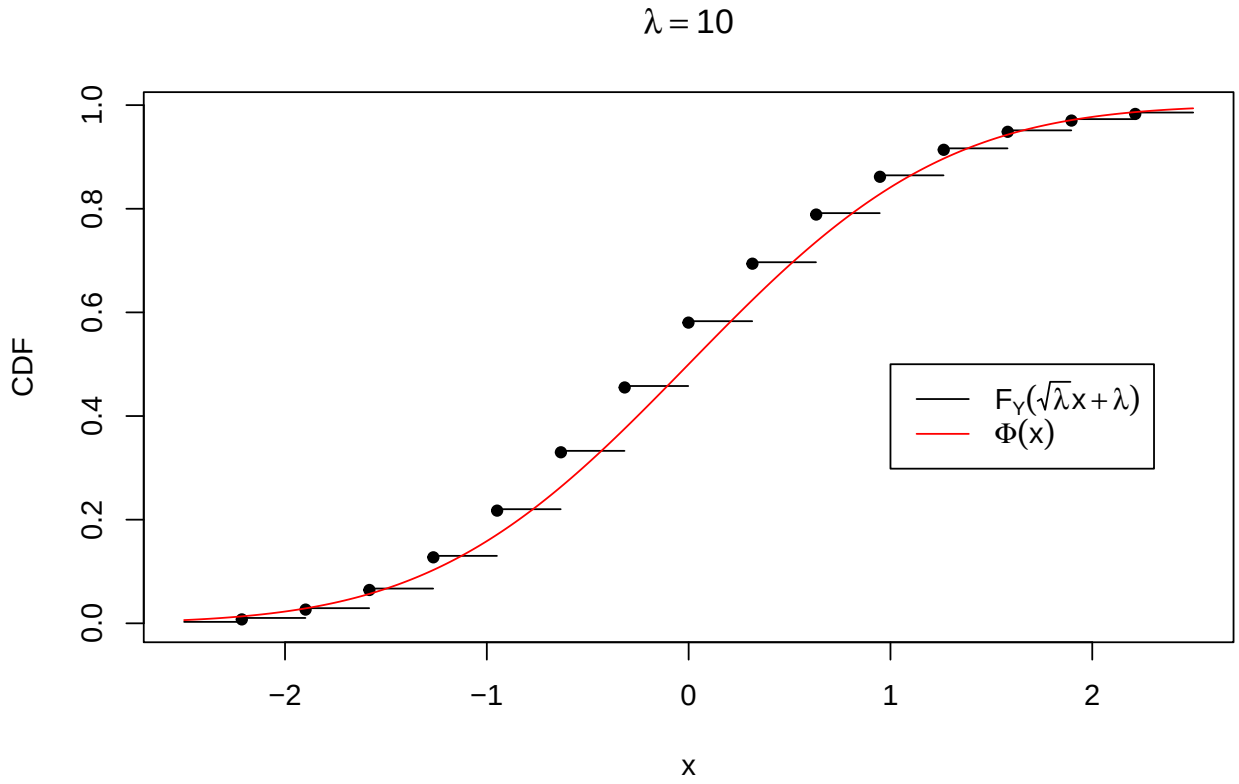
$$F_X(x) \doteq F_Y\left(\sqrt{\lambda}x + \lambda\right)$$

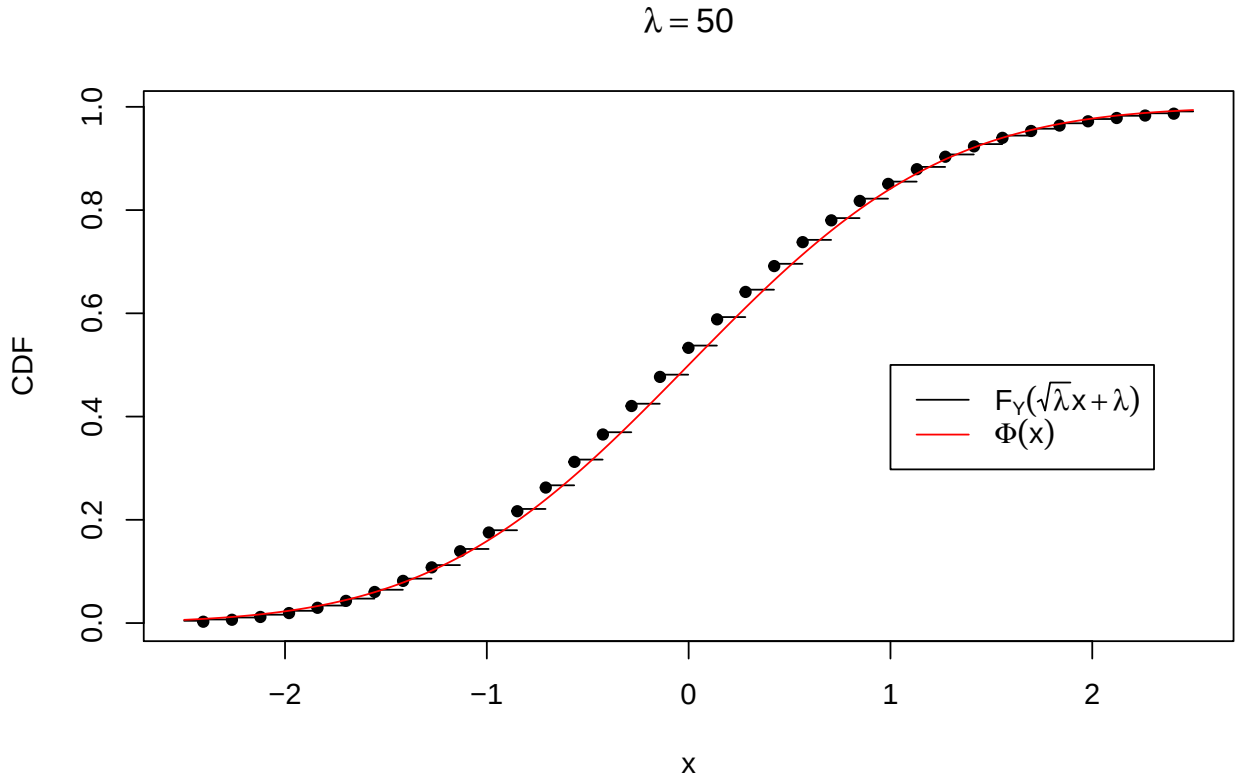
using numerical and graphical means. We can compute the exact Poisson cdf using the `R` function `ppois`, and the Normal approximation using `pnorm`. Note that the notation $\Phi(x)$ indicates the standard Normal cdf, that is

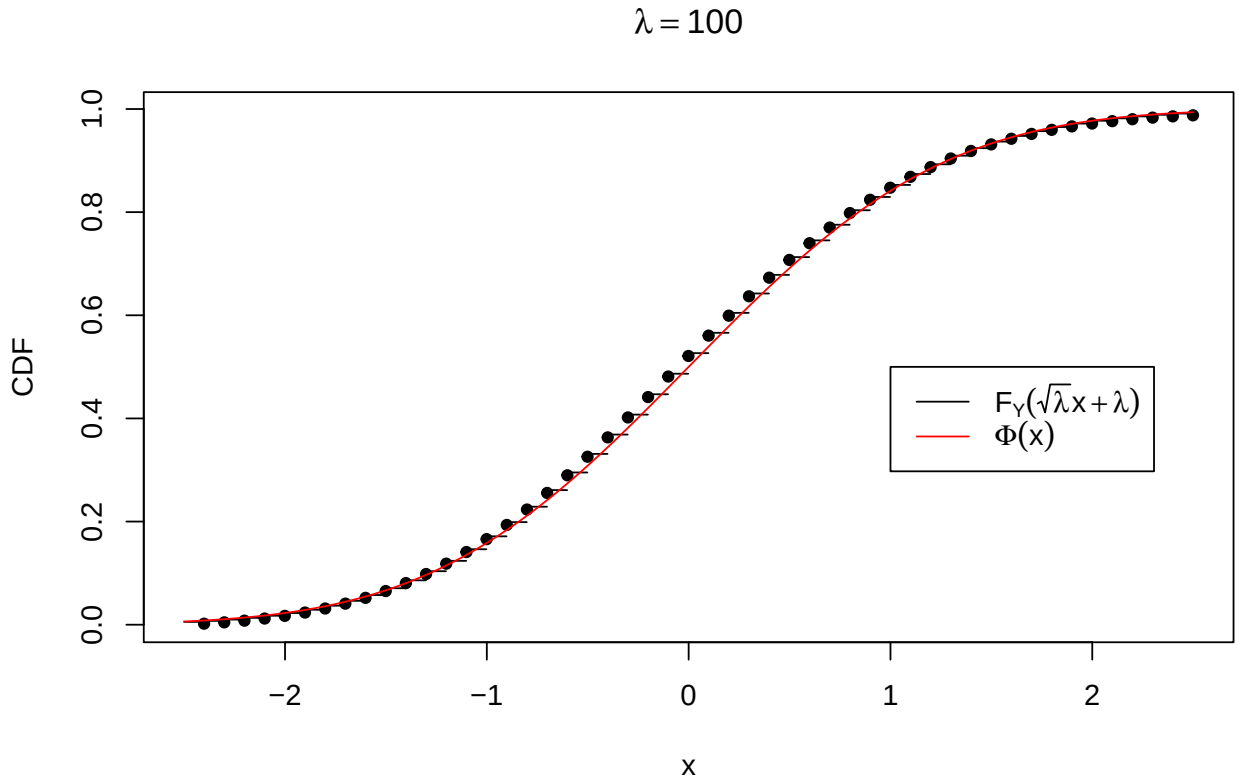
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du.$$

```
set.seed(1010)
x<-seq(-2.5,2.5,by=0.001)
lvec<-c(5,10,50,100)
for(lam in lvec){
    y<-sqrt(lam)*x+lam
    Fx<-ppois(y,lam)
    px<-diff(Fx)
    Fx[px > 0]<-NA
    plot(x,Fx,pch=19,type='l',ylab='CDF')
    points(x[px > 0],cumsum(px[px>0]),pch=19,cex=0.8)
    lines(x,pnorm(x),col='red')
    title(substitute(lambda==l,list(l=lam)))
    legend(1,0.5,c(expression(F[Y](sqrt(lambda)*x+lambda)),
                expression(Phi(x))),col=c('black','red'),lty=1)
}
```


λ = 5
CDF
0.0
0.2
0.4
0.6
0.8
1.0
−2
−1
0
1
2
x
$F_Y(\sqrt{\lambda}x+\lambda)$
Φ(x)


λ = 10
CDF
0.0
0.2
0.4
0.6
0.8
1.0
−2
−1
0
1
2
x
$F_Y(\sqrt{\lambda}x+\lambda)$
Φ(x)





Clearly the approximation is good.

We can also use this method to approximate the pmf, by noting that for $y = 1, 2, \ldots$, we have

$$p_Y(y) = \begin{cases} F_Y(0) & y = 0 \\ F_Y(y) - F_Y(y-1) & y = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

and also that

$$F_Y(y) = F_X\left(\frac{y-\lambda}{\sqrt{\lambda}}\right).$$

In this application, the *continuity correction* can be useful: we instead approximate

$$F_Y(y) = F_X\left(\frac{y+\frac{1}{2}-\lambda}{\sqrt{\lambda}}\right).$$

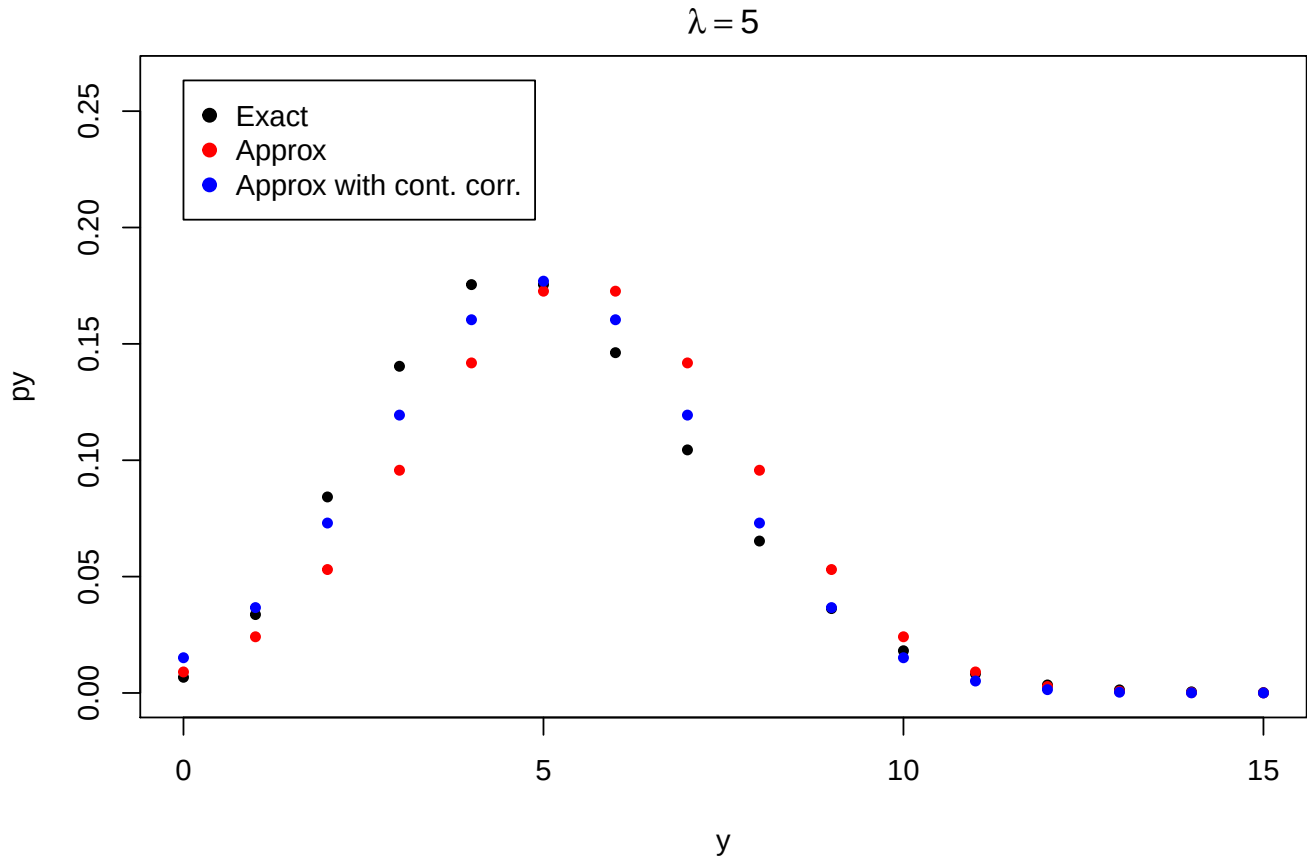
The continuity correction is a way to account for the fact that we are approximating the discrete distribution F_Y by the continuous Normal approximation; the integer value y can be regarded as a 'rounded down' version of a continuous underlying variable.

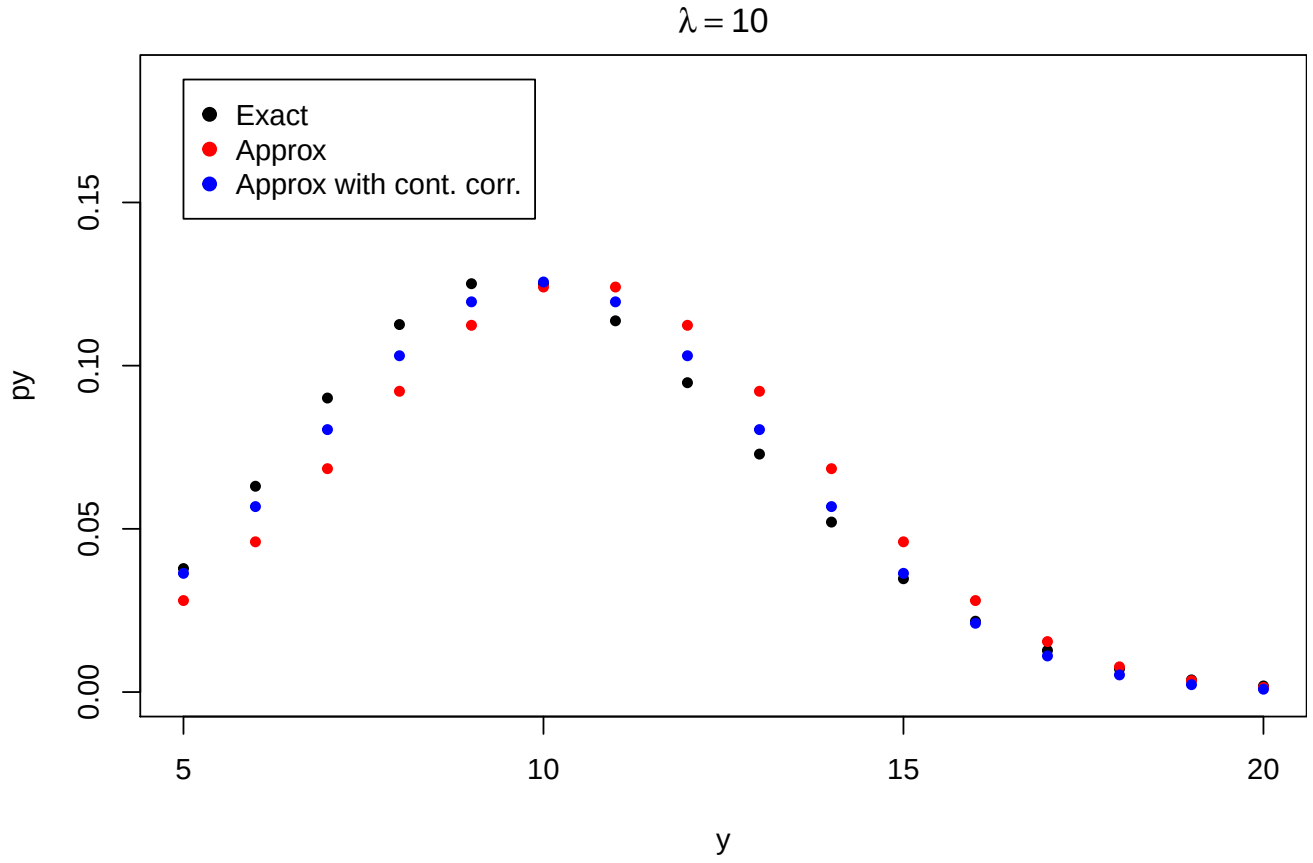
The approximation is most accurate when y is near λ, so the following plots compute

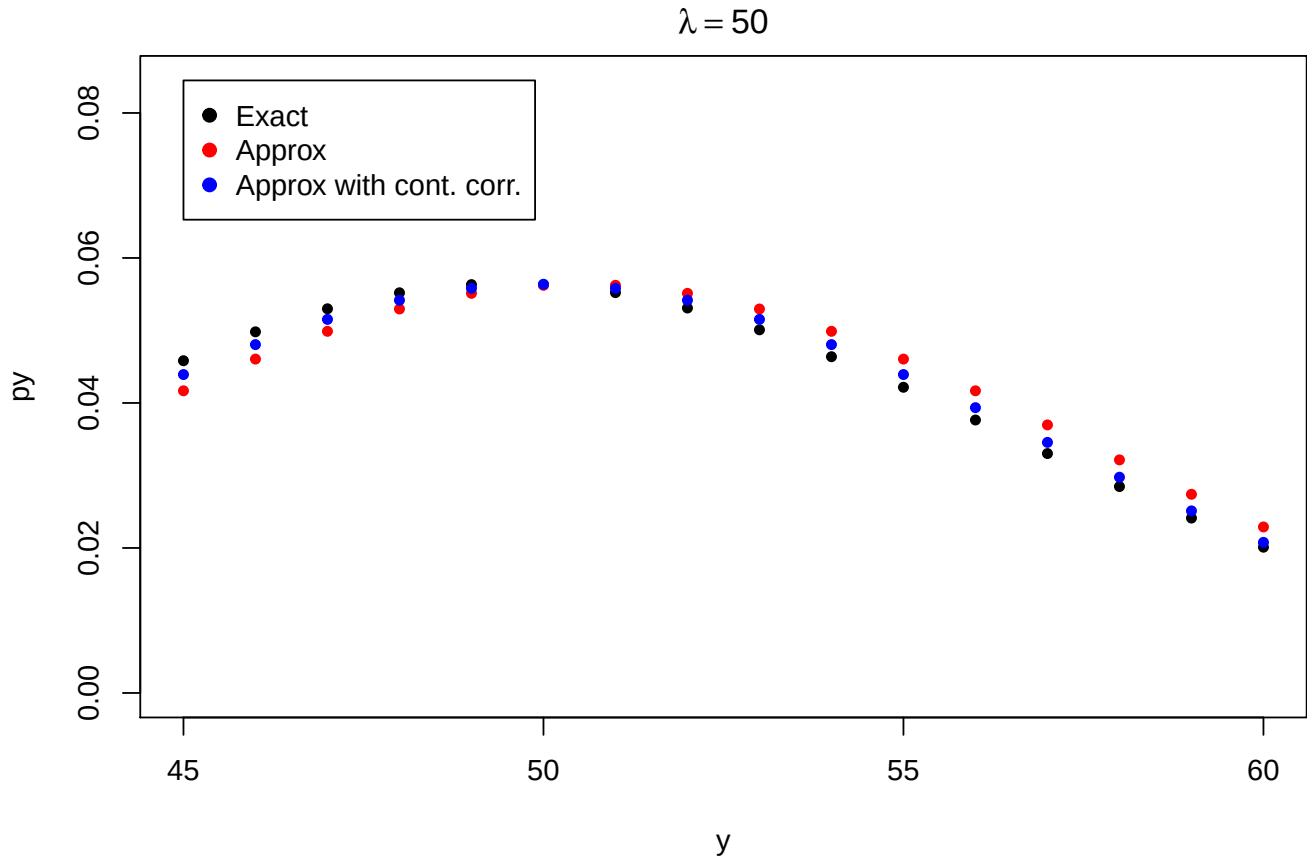
- the exact pmf (black)
- the approximation to the pmf using the Normal distribution (red)
- the approximation to the pmf using the Normal distribution and a continuity correction (blue)

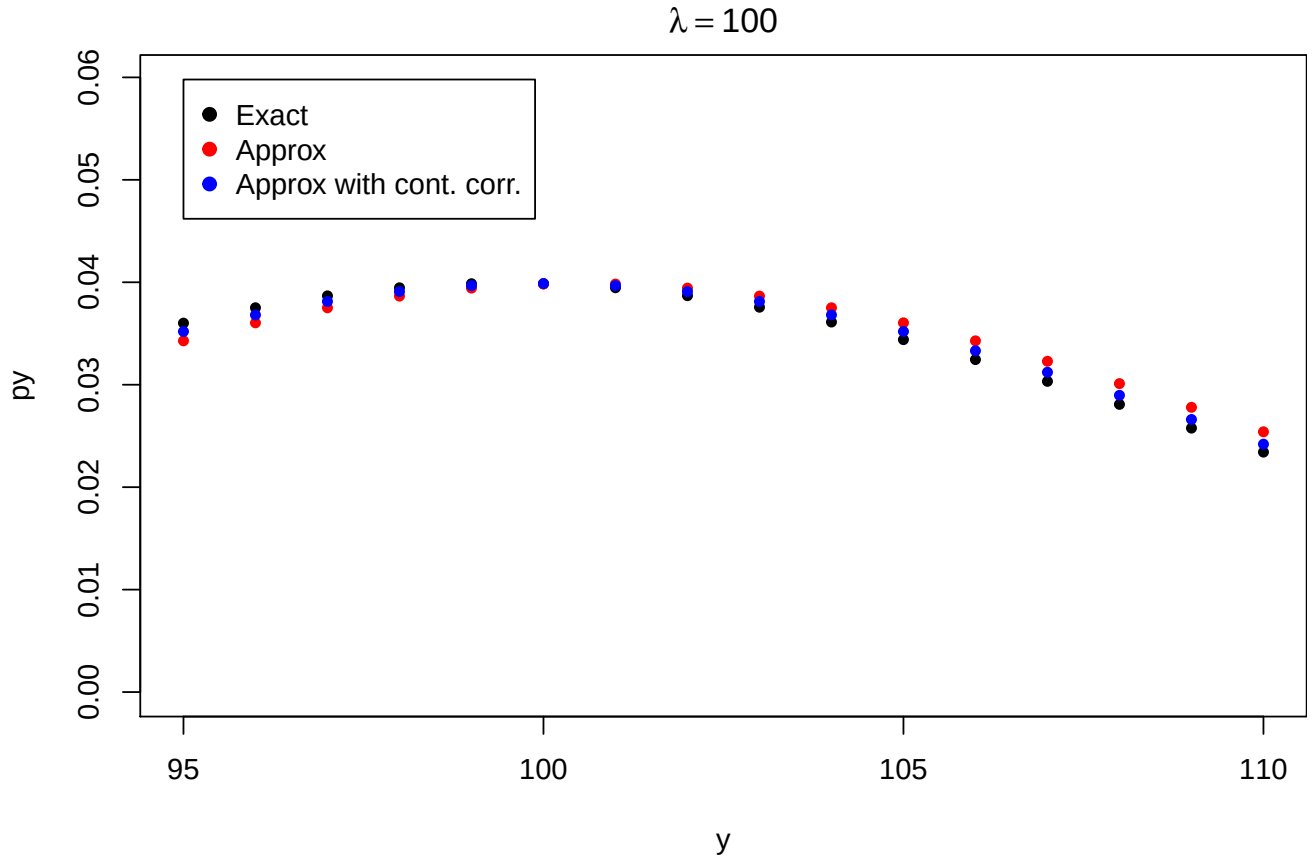
with all pmfs being computed in the range $\lambda - 5, \lambda - 4, \ldots, \lambda + 9, \lambda + 10$, for $\lambda = 5, 10, 50, 100, 200, 500$.

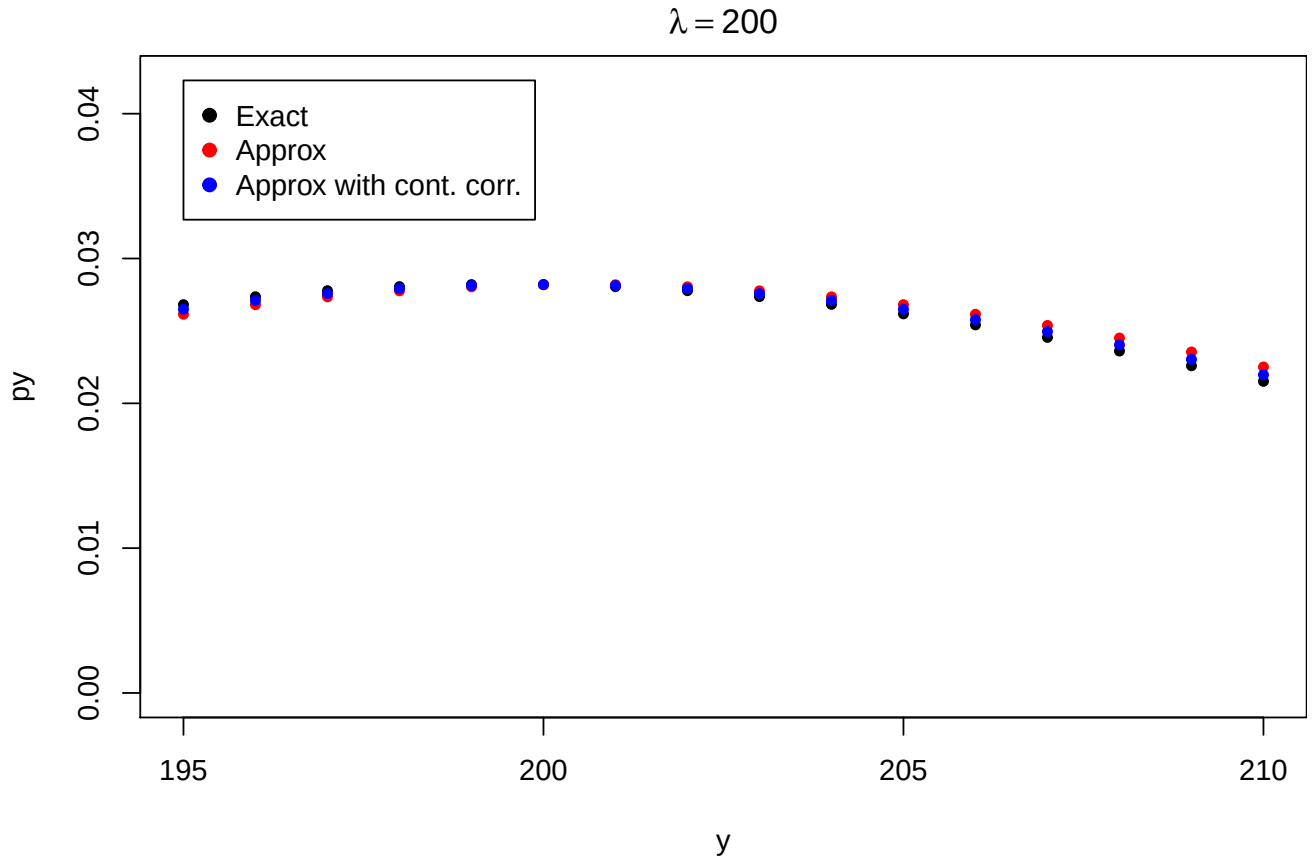
```
lvec<-c(5,10,50,100,200,500)
for(lam in lvec){
    y<-seq(lam-6,lam+10,by=1)
    py<-dpois(y,lam)[-1]
    pmax<-1.5*max(py)
    Fy.approx<-pnorm((y-lam)/sqrt(lam))
    py.approx<-diff(Fy.approx)
    Fy.approxCC<-pnorm((y+0.5-lam)/sqrt(lam))
    py.approxCC<-diff(Fy.approxCC)
    par(mar=c(4,4,2,0))
    plot(y[-1],py,pch=19,ylim=range(0,pmax),cex=0.7,xlab='y')
    points(y[-1],py.approx,col='red',pch=19,cex=0.7)
    points(y[-1],py.approxCC,col='blue',pch=19,cex=0.7)
    title(substitute(lambda==l,list(l=lam)))
    legend(y[2],pmax,c('Exact','Approx','Approx with cont. corr.'),
        col=c('black','red','blue'),pch=19)
}
```

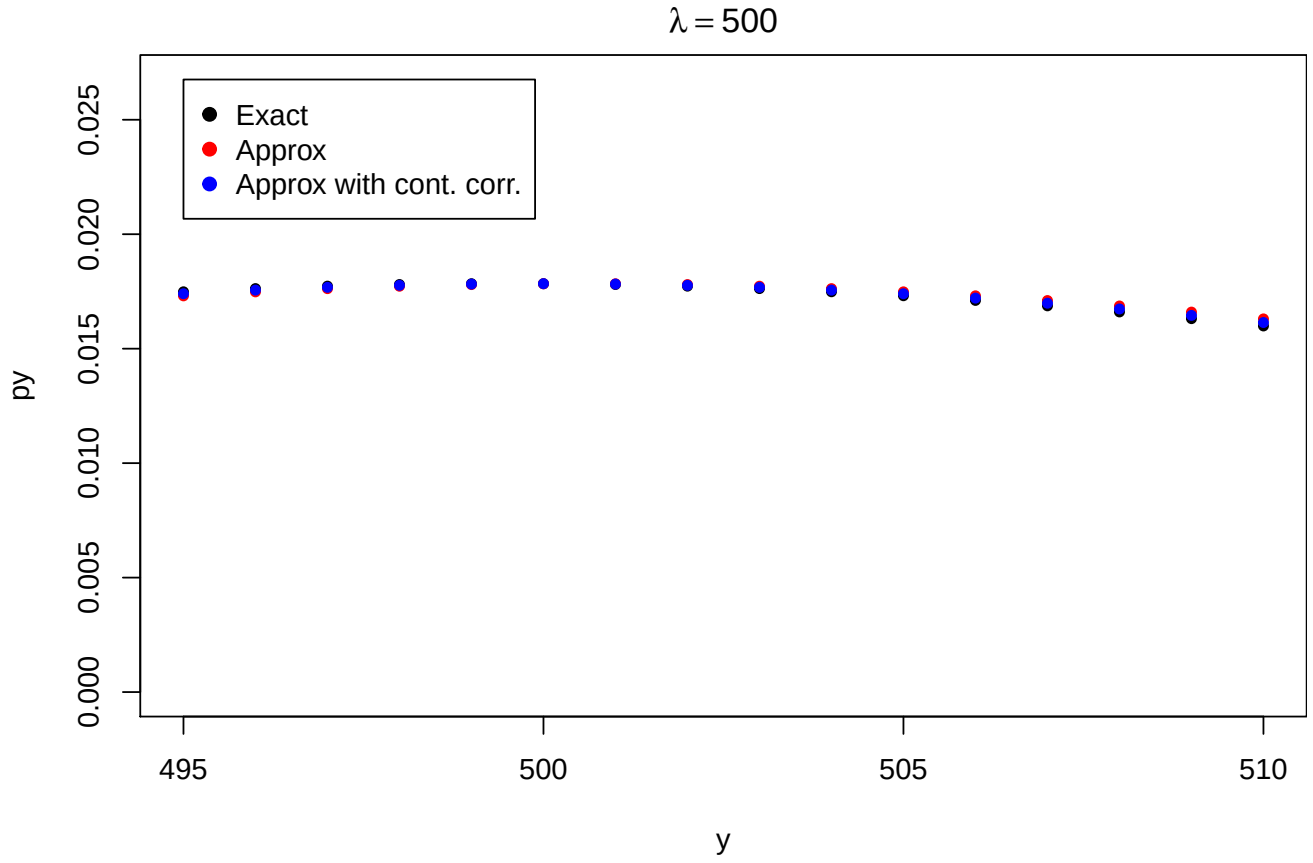

λ = 5
Exact
Approx
Approx with cont. corr.
py
0.00
0.05
0.10
0.15
0.20
0.25
0
5
10
15
y


λ = 10
Exact
Approx
Approx with cont. corr.
py
0.00
0.05
0.10
0.15
5
10
15
20
y


λ = 50
Exact
Approx
Approx with cont. corr.
py
0.00
0.02
0.04
0.06
0.08
45
50
55
60
y


λ = 100
Exact
Approx
Approx with cont. corr.
py
0.00
0.01
0.02
0.03
0.04
0.05
0.06
95
100
105
110
y


λ = 200
Exact
Approx
Approx with cont. corr.
py
0.00
0.01
0.02
0.03
0.04
195
200
205
210
y


λ = 500
Exact
Approx
Approx with cont. corr.
py
0.000
0.005
0.010
0.015
0.020
0.025
495
500
505
510
y