Oblig 2 Statistikk

---

1 Kapittel 7

1.1 a) Kan du bruke P(X <= a)for å finne P(X > a)? Hvordan?

Ja man kan finne det ut ved å ta 1-P(X=<) fordi da får vi den resterende sannsynligheten. Vi vet sannsynligheten for mindre enn a så det som er igjen blir større enn a.

1.2 b) Hvorfor er P(X < c) = P(X <= c) når X er kontinuerlig?

Fordi når X er kontinuerlig så kan aldri sannsynligheten være eksakt på et punkt. Dermed er sannsynligheten mindre enn c lik som når vi tar med akkurat punktet c.

1.3 c) Hvorfor kan vi ikke regne med P(X < c) = P(X <= c) når X er diskret? (Hvorfor vil de for det meste være forskjellige?)

Fordi når X er diskret så kan vi ha en sannsynlighet i punktet c som ikke er lik den totale sannsynligheten mindre enn c.

1.4 d) For hånd: 1e (en: 2e)

1.5 e) For hånd: 2b (en: 3b)

1.6 f) Gjør i R: 1e (en: 2e)

1.7 g) Gjør i R: 2b (en: 3b)

2 Kapittel 9

2.1 a) X bin(20,0.375). Lag tabell over sannsynligheter for x= 0,....,20, og plott både pdf og CDF for denne sannsynlighetsfordelingen.

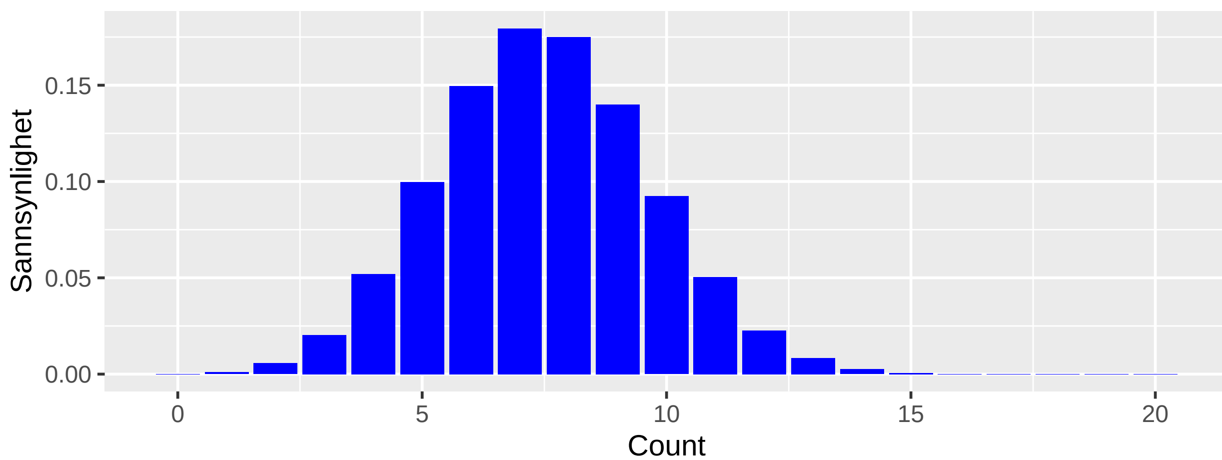
2.1.1 Her er tabell og plot for PDF:

```
library(ggplot2)

x = c(0:20)

y = dbinom(x, 20, 0.375)

df = data.frame(x, y)

df
```

```
##     x            y
## 1   0 8.271806e-05
## 2   1 9.926167e-04
```

```
## 3   2 5.657915e-03
## 4   3 2.036850e-02
## 5   4 5.193966e-02
## 6   5 9.972415e-02
## 7   6 1.495862e-01
## 8   7 1.795035e-01
## 9   8 1.750159e-01
## 10  9 1.400127e-01
## 11 10 9.240839e-02
## 12 11 5.040458e-02
## 13 12 2.268206e-02
## 14 13 8.374914e-03
## 15 14 2.512474e-03
## 16 15 6.029938e-04
## 17 16 1.130613e-04
## 18 17 1.596160e-05
## 19 18 1.596160e-06
## 20 19 1.008101e-07
## 21 20 3.024303e-09
```

```
ggplot(data=df, aes(x=x, y=y)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```



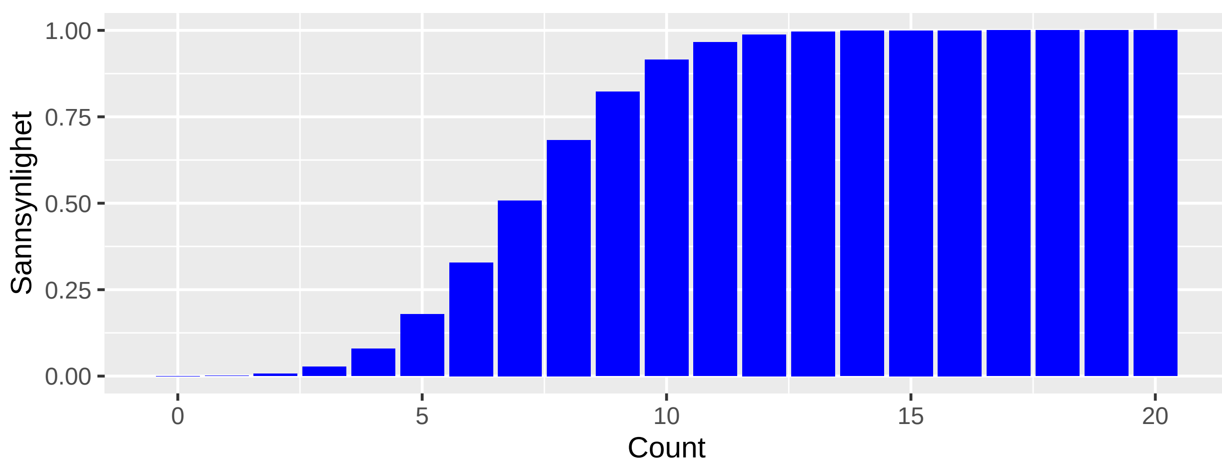
2.1.2 Her er tabell og plot for CDF:

```
y_cdf = pbinom(x, 20, 0.375)

df_cdf = data.frame(x, y_cdf)

df_cdf
```

```
##      x        y_cdf
## 1    0 8.271806e-05
## 2    1 1.075335e-03
## 3    2 6.733250e-03
## 4    3 2.710175e-02
## 5    4 7.904141e-02
## 6    5 1.787656e-01
```

```
## 7    6 3.283518e-01
## 8    7 5.078553e-01
## 9    8 6.828712e-01
## 10   9 8.228839e-01
## 11  10 9.152923e-01
## 12  11 9.656968e-01
## 13  12 9.883789e-01
## 14  13 9.967538e-01
## 15  14 9.992663e-01
## 16  15 9.998693e-01
## 17  16 9.999823e-01
## 18  17 9.999983e-01
## 19  18 9.999999e-01
## 20  19 1.000000e+00
## 21  20 1.000000e+00
```

```
ggplot(data=df_cdf, aes(x=x, y=y_cdf)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```



2.1.3 E[X]

Vi summerer opp hver count ganget med sannsynligheten for å finne E[X]

```
sum(x*0.375)
```

E[X] = 78.75

2.1.4 Var(X)

For å finne var(X) kjører vi bare følgende R-kode

```
var(x)
```

Var(X) = 38.5

2.1.5 P(2 < X < 7)

For å finne ut sannsynligheten for X mellom 2 og 7 skriver vi følgende R-kode

```
lessthan7 = pbinom(6, length(x), 0.375)
lessthan2 = pbinom(1, length(x), 0.375)
```

```r
print(lessthan7-lessthan2)
```

```
## [1] 0.2715539
```

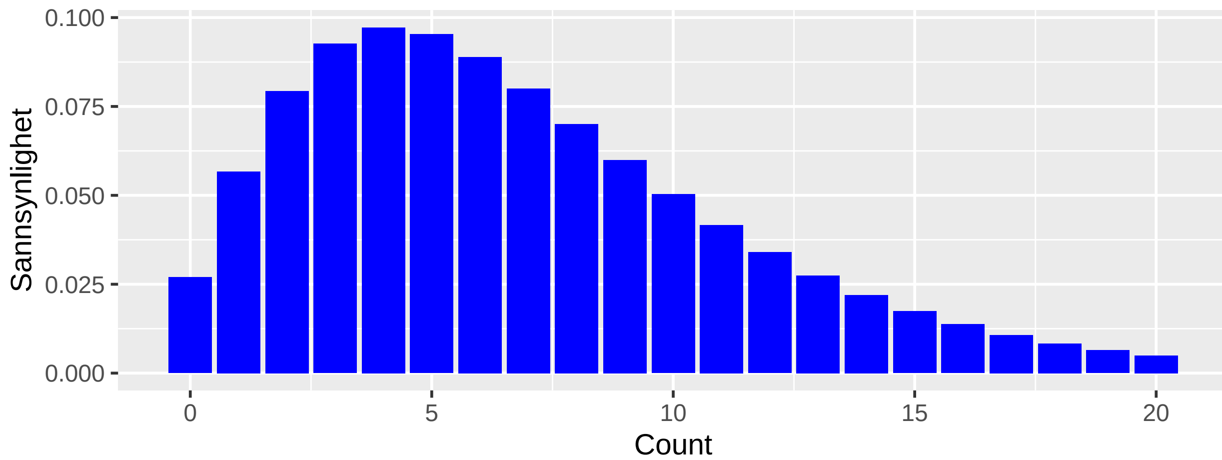
2.2 b) X nb(3,0.3). Lag tabell over sannsynligheter for x = 0,....,20, og plott både pdf og CDF for denne sannsynlighetsfordelingen

2.2.1 Her er tabell og plot for PDF:

```r
x = c(0:20)

y = dnbinom(x, 3, 0.3)

nbdf_pdf = data.frame(x, y)

nbdf_pdf
```

```
##     x           y
## 1   0 0.027000000
## 2   1 0.056700000
## 3   2 0.079380000
## 4   3 0.092610000
## 5   4 0.097240500
## 6   5 0.095295690
## 7   6 0.088942644
## 8   7 0.080048380
## 9   8 0.070042332
## 10  9 0.059925106
## 11 10 0.050337089
## 12 11 0.041642501
## 13 12 0.034008043
## 14 13 0.027468034
## 15 14 0.021974428
## 16 15 0.017433046
## 17 16 0.013728524
## 18 17 0.010740551
## 19 18 0.008353762
## 20 19 0.006463174
## 21 20 0.004976644
```

```r
ggplot(data=nbdf_pdf, aes(x=x, y=y)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```

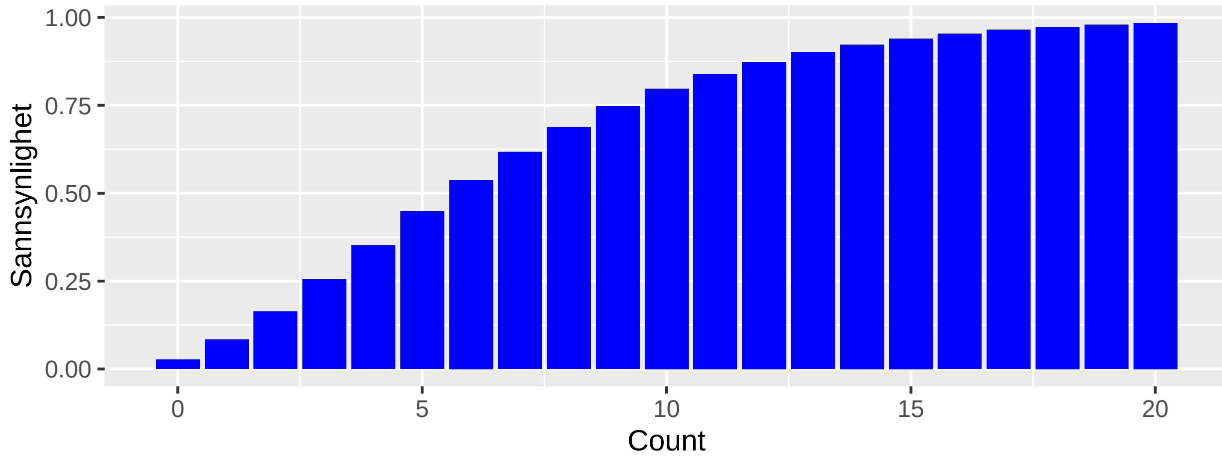


2.2.2 Her er tabell og plot for CDF:

```
nby_cdf = pnbinom(x, 3, 0.3)

nbdf_cdf = data.frame(x, nby_cdf)

nbdf_cdf
```

```
##     x   nby_cdf
## 1   0 0.0270000
## 2   1 0.0837000
## 3   2 0.1630800
## 4   3 0.2556900
## 5   4 0.3529305
## 6   5 0.4482262
## 7   6 0.5371688
## 8   7 0.6172172
## 9   8 0.6872595
## 10  9 0.7471847
## 11 10 0.7975217
## 12 11 0.8391642
## 13 12 0.8731723
## 14 13 0.9006403
## 15 14 0.9226147
## 16 15 0.9400478
## 17 16 0.9537763
## 18 17 0.9645169
## 19 18 0.9728706
## 20 19 0.9793338
## 21 20 0.9843104
```

```
ggplot(data=nbdf_cdf, aes(x=x, y=nby_cdf)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```



2.2.3 E[X]

Vi summerer opp hver count ganget med sannsynligheten for å finne E[X]

```
sum(x*0.3)
```

E[X] = 63

2.2.4 Var(X)

For å finne var(X) kjører vi bare følgende R-kode

```
var(x)
```

Var(X) = 38.5

2.2.5 P(2 < X < 7)

For å finne ut sannsynligheten for X mellom 2 og 7 skriver vi følgende R-kode

```
lessthan7 = pnbinom(6, 3, 0.3)
lessthan2 = pnbinom(1, 3, 0.3)

print(lessthan7-lessthan2)
```

```
## [1] 0.4534688
```

2.3 X pois7.8. Lag tabell over sannsynligheter forx= 0,....,20, og plott både pdf og CDF for denne sannsynlighetsfordelingen

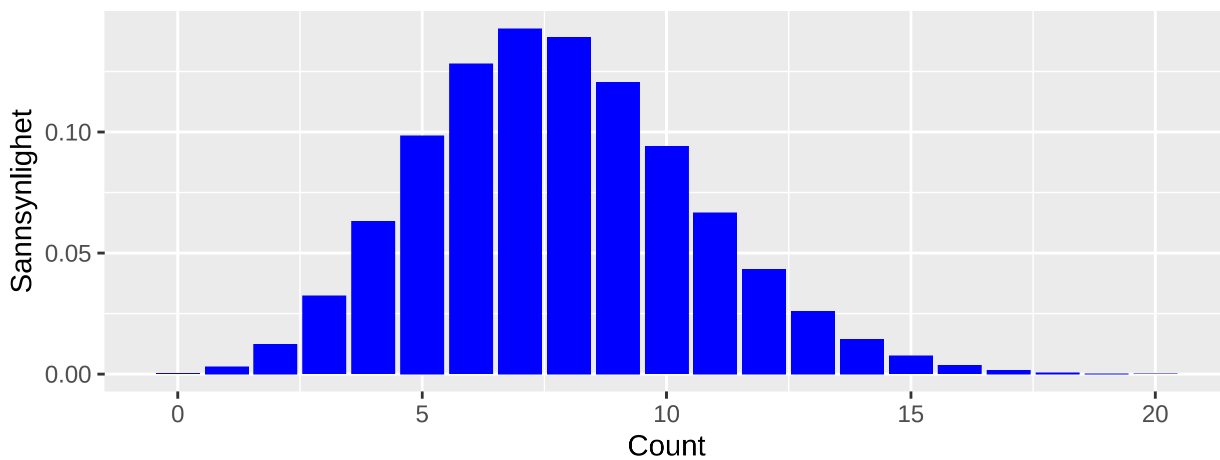
2.3.1 Her er tabell og plot for PDF:

```
py_pdf = dpois(x, 7.8)

p_df_pdf = data.frame(x, py_pdf)

p_df_pdf
```

```
##     x       py_pdf
## 1   0 0.0004097350
## 2   1 0.0031959328
```

```
## 3    2 0.0124641381
## 4    3 0.0324067590
## 5    4 0.0631931800
## 6    5 0.0985813607
## 7    6 0.1281557690
## 8    7 0.1428021426
## 9    8 0.1392320890
## 10   9 0.1206678105
## 11  10 0.0941208922
## 12  11 0.0667402690
## 13  12 0.0433811748
## 14  13 0.0260287049
## 15  14 0.0145017070
## 16  15 0.0075408877
## 17  16 0.0036761827
## 18  17 0.0016867191
## 19  18 0.0007309116
## 20  19 0.0003000585
## 21  20 0.0001170228
```

```r
ggplot(data=p_df_pdf, aes(x=x, y=py_pdf)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```



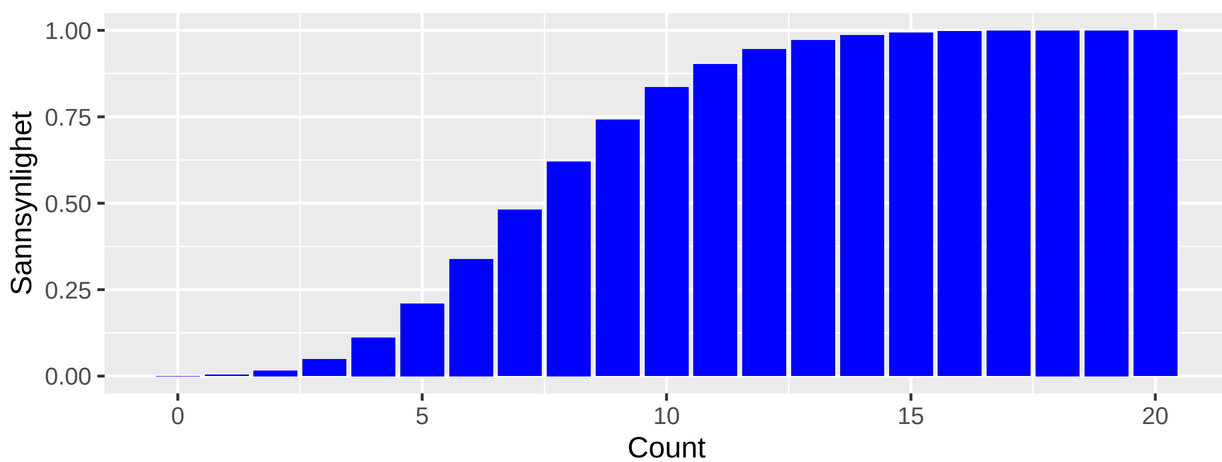
2.3.2 Her er tabell og plot for CDF:

```r
pby_cdf = ppois(x, 7.8)

pdf_cdf = data.frame(x, nby_cdf)

pdf_cdf
```

```
##      x   nby_cdf
## 1    0 0.0270000
## 2    1 0.0837000
## 3    2 0.1630800
## 4    3 0.2556900
## 5    4 0.3529305
## 6    5 0.4482262
```

```
## 7    6 0.5371688
## 8    7 0.6172172
## 9    8 0.6872595
## 10   9 0.7471847
## 11  10 0.7975217
## 12  11 0.8391642
## 13  12 0.8731723
## 14  13 0.9006403
## 15  14 0.9226147
## 16  15 0.9400478
## 17  16 0.9537763
## 18  17 0.9645169
## 19  18 0.9728706
## 20  19 0.9793338
## 21  20 0.9843104
```

```r
ggplot(data=pdf_cdf, aes(x=x, y=pby_cdf)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```



2.3.3 E[X]

Vi summerer opp hver count ganget med sannsynligheten for å finne E[X]

```r
sum(x*7.8)
```

E[X] = 1638

2.3.4 Var(X)

For å finne var(X) kjører vi bare følgende R-kode

```r
var(x)
```

Var(X) = 38.5

2.3.5 $P(2 < X < 7)$

For å finne ut sannsynligheten for X mellom 2 og 7 skriver vi følgende R-kode

```r
lessthan7 = ppois(6, 7.8)
lessthan2 = ppois(1, 7.8)
```

```r
print(lessthan7-lessthan2)
```

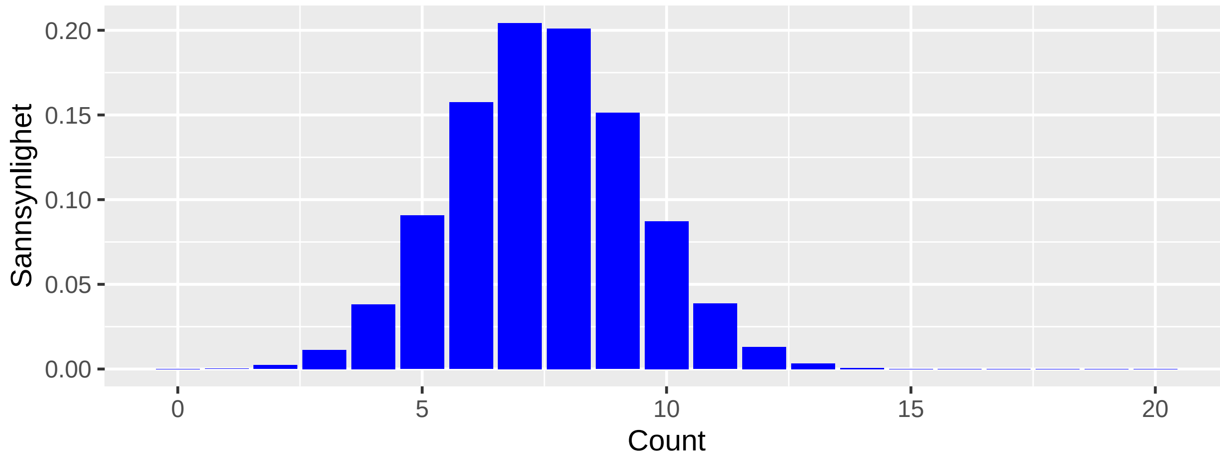
```
## [1] 0.3348012
```

2.4 Hypergeometrisk:X hyp(20,30,80). Lag tabell over sannsynligheter forx=0,....,20, og plott både pdf og CDF for denne sannsynlighetsfordelingen.

```r
hypery_pdf = dhyper(x, 30, 50, 20)

hyperydf_pdf = data.frame(x, hypery_pdf)

hyperydf_pdf
```

```
##     x   hypery_pdf
## 1   0 1.333098e-05
## 2   1 2.580189e-04
## 3   2 2.221381e-03
## 4   3 1.130885e-02
## 5   4 3.816737e-02
## 6   5 9.072929e-02
## 7   6 1.575161e-01
## 8   7 2.043452e-01
## 9   8 2.009843e-01
## 10  9 1.511677e-01
## 11 10 8.729934e-02
## 12 11 3.871367e-02
## 13 12 1.313500e-02
## 14 13 3.383613e-03
## 15 14 6.536525e-04
## 16 15 9.296391e-05
## 17 16 9.473224e-06
## 18 17 6.639556e-07
## 19 18 2.997022e-08
## 20 19 7.725943e-10
## 21 20 8.498537e-12
```

```r
ggplot(data=hyperydf_pdf, aes(x=x, y=hypery_pdf)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```

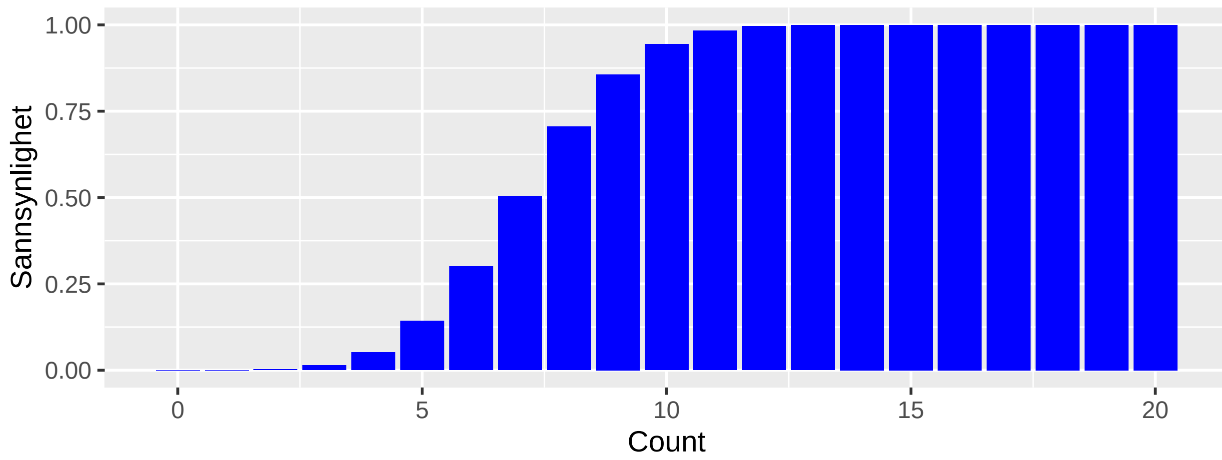


2.4.1 Her er tabell og plot for CDF:

```
hypery_cdf = phyper(x, 30,50,20)

hyperydf_cdf = data.frame(x, hypery_cdf)

hyperydf_cdf
```

```
##      x   hypery_cdf
## 1    0 1.333098e-05
## 2    1 2.713499e-04
## 3    2 2.492731e-03
## 4    3 1.380158e-02
## 5    4 5.196895e-02
## 6    5 1.426982e-01
## 7    6 3.002144e-01
## 8    7 5.045596e-01
## 9    8 7.055439e-01
## 10   9 8.567116e-01
## 11  10 9.440109e-01
## 12  11 9.827246e-01
## 13  12 9.958596e-01
## 14  13 9.992432e-01
## 15  14 9.998969e-01
## 16  15 9.999898e-01
## 17  16 9.999993e-01
## 18  17 1.000000e+00
## 19  18 1.000000e+00
## 20  19 1.000000e+00
## 21  20 1.000000e+00
```

```
ggplot(data=hyperydf_cdf, aes(x=x, y=hypery_cdf)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x = "Count", y = "Sannsynlighet")
```



2.4.2 E[X]

Vi summerer opp hver count ganget med sannsynligheten for å finne E[X]

```
sum(x*hypery_pdf)
```

E[X] = 7.5

2.4.3 Var(X)

For å finne var(X) kjører vi bare følgende R-kode

```
p=30/80
var_x = 20*p*(1 - p)*(30+50-20)/(30+50-1)
```

Var(X) = 38.5

2.4.4 $P(2 < X < 7)$

For å finne ut sannsynligheten for X mellom 2 og 7 skriver vi følgende R-kode

```
lessthan7 = phyper(6,30,50,20)
lessthan2 = phyper(1,30,50,20)

print(lessthan7-lessthan2)
```

```
## [1] 0.299943
```

2.5 X Betab(3a,5a,20)for $a = 2$. Lag tabell over sannsynligheter for $x = 0,....,20$, og regn deretter ut E[X],Var(X), og$P(2 < X < 7)$. Plott både pdf og CDF for disse sannsynlighetsfordelingene, med$a = 1$,$a = 2$,$a = 4$, og $a = 10$, og sammenlign medtilsvarende plott for bin(20,0.375)