P8130 Recitation 2: Sept 25th/27th

Yutao Liu

September 15, 2017

Key Words: Convergence in Distribution; Monte Carlo Simulation

```
rm( list = ls() ) # clear workspace

if ( !require(pacman) ) install.packages('pacman')
pacman::p_load(dplyr, ggplot2)
pacman::p_load(gridExtra)
```

1) Poisson Approximation to the Binomial Distribution (Rosner, Chapter 4.13; Casella & Berger, Example 2.3.13)

$\text{Bin}(n = 100, p = .935)$

```
n.rep <- 1e5

n <- 100; p <- .935; n*p
```

```
## [1] 93.5
```

```
set.seed(2017)
bin.rv1 <- rbinom(n.rep, n, p)
pois.rv1 <- rpois(n.rep, lambda = n*p)
```

Why set seed? Reproducibility!!

```
rbinom(n = 10, size = 15, p = .3)
```

```
##  [1] 8 4 2 5 5 2 5 3 3 5
```

```
rbinom(n = 10, size = 15, p = .3)
```

```
##  [1] 5 2 5 6 3 7 6 5 6 4
```

```
set.seed(1); rbinom(n = 10, size = 15, p = .3)
```

```
##  [1] 3 4 5 7 3 7 7 5 5 2
```

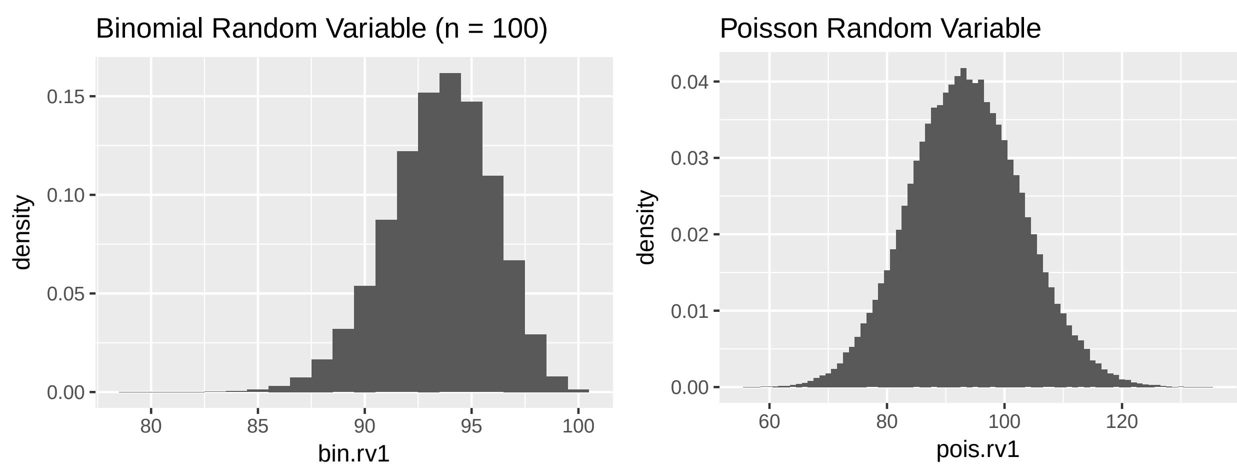
```
set.seed(1); rbinom(n = 10, size = 15, p = .3)
```

```
##  [1] 3 4 5 7 3 7 7 5 5 2
```

```
set.seed(2017); rbinom(n = 10, size = 15, p = .3)
```

```
##  [1] 7 5 4 3 6 6 2 4 4 3
```

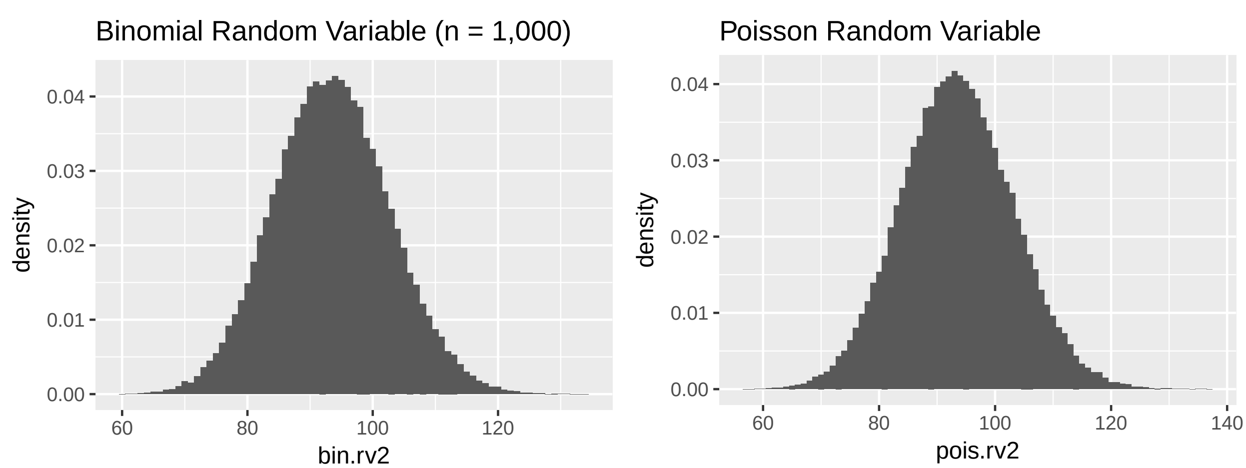
```
p1 <- ggplot(mapping = aes(bin.rv1) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
        labs(title = "Binomial Random Variable (n = 100)")
p2 <- ggplot(mapping = aes(pois.rv1) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
        labs(title = "Poisson Random Variable")
grid.arrange(p1, p2, ncol = 2)
```




```r
n <- 1e3; p <- 9.35e-2; n*p
```

```
## [1] 93.5
```

```r
set.seed(2017)
bin.rv2 <- rbinom(n.rep, n, p)
pois.rv2 <- rpois(n.rep, lambda = n*p)

p3 <- ggplot(mapping = aes(bin.rv2) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
        labs(title = "Binomial Random Variable (n = 1,000)")
p4 <- ggplot(mapping = aes(pois.rv2) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
        labs(title = "Poisson Random Variable")
grid.arrange(p3, p4, ncol = 2)
```

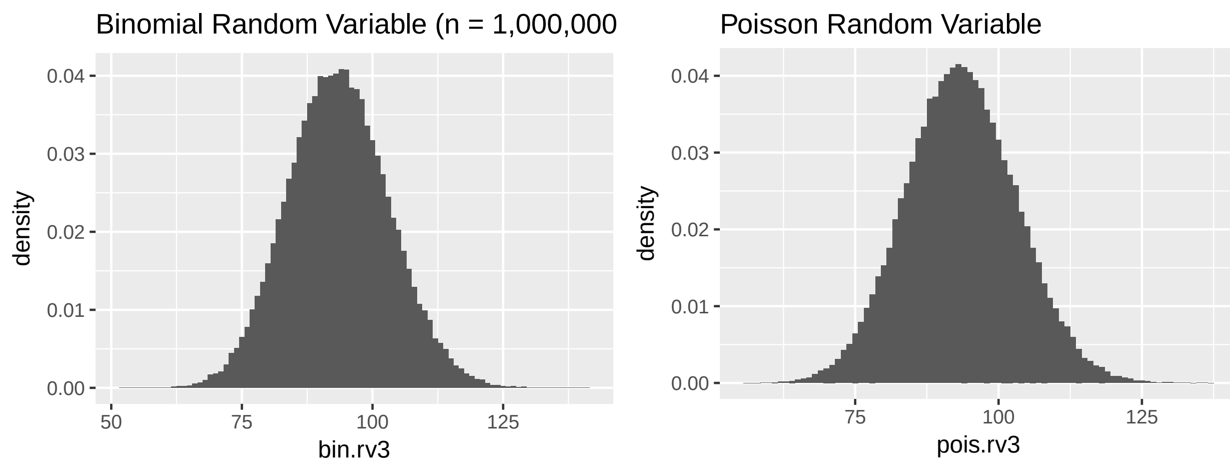



```r
n <- 1e6; p <- 9.35e-5; n*p
```

```
## [1] 93.5
```

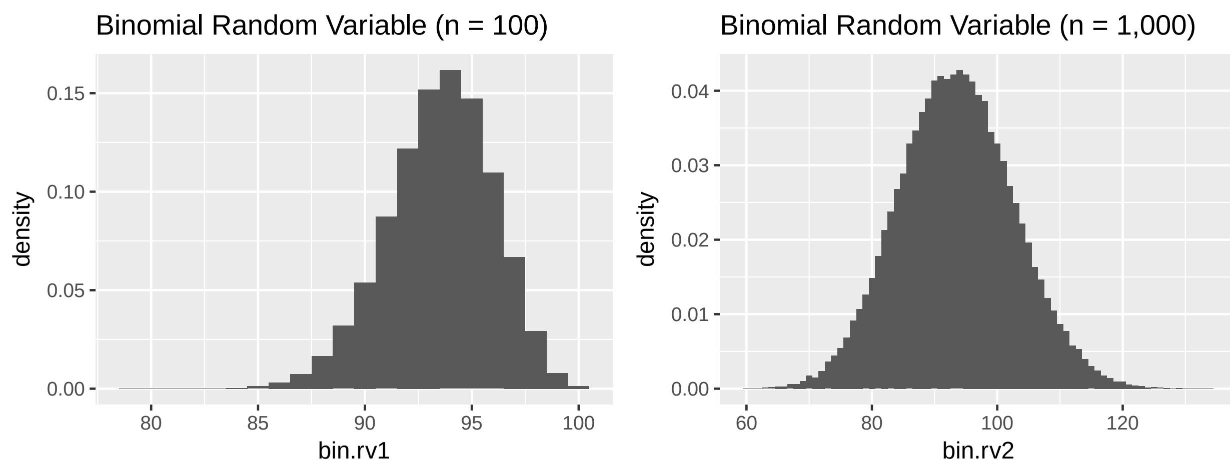
```r
set.seed(2017)
bin.rv3 <- rbinom(n.rep, n, p)
pois.rv3 <- rpois(n.rep, lambda = n*p)

p5 <- ggplot(mapping = aes(bin.rv3) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
        labs(title = "Binomial Random Variable (n = 1,000,000")
p6 <- ggplot(mapping = aes(pois.rv3) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
```

```r
        labs(title = "Poisson Random Variable")
grid.arrange(p5, p6, ncol = 2)
```

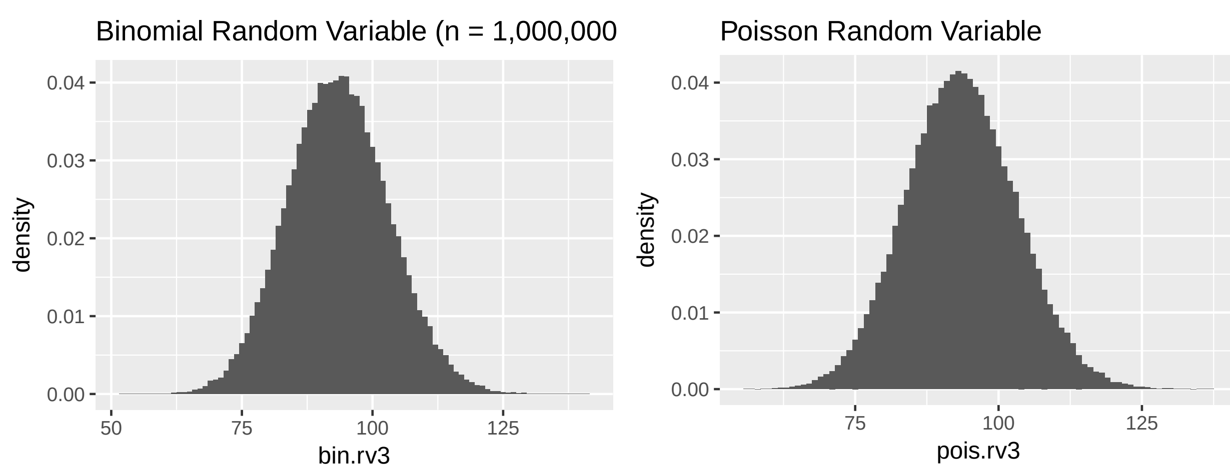


Convergence

```r
grid.arrange(p1, p3, ncol = 2)
```




```r
grid.arrange(p5, p6, ncol = 2)
```



2) Central Limit Theorem (Rosner, Equation 6.3; Casella & Berger, Example 2.3.13)

(Rosner, Equation 6.3) Let $X_1, \ldots, X_n$ be a random sample from population with mean μ and variance σ^2. Then for large n, $\bar{X} \dot{\sim} N(\mu, \sigma^2/n)$ even if the underlying distribution of individual observations in the population is not normal.

Example 1: Sample Mean of a Binary Sample

$X \sim Bin(n = 1, p = .3)$

```
x <- rbinom(n.rep, size = 1, prob = .3)
p7 <- ggplot(mapping = aes(x) ) + geom_histogram(aes(y=..density..), binwidth = 1) +
        labs(title = "Binary Random Variable")

prop <- function(smpl.size, prob, print.smpl = FALSE) {
  smpl <- rbinom(n = smpl.size, size = 1, prob)
  if (print.smpl) print(smpl)
  mean(smpl) %>% return(.)
}

prop (smpl.size = 20, prob = .3, print.smpl = TRUE)
```

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

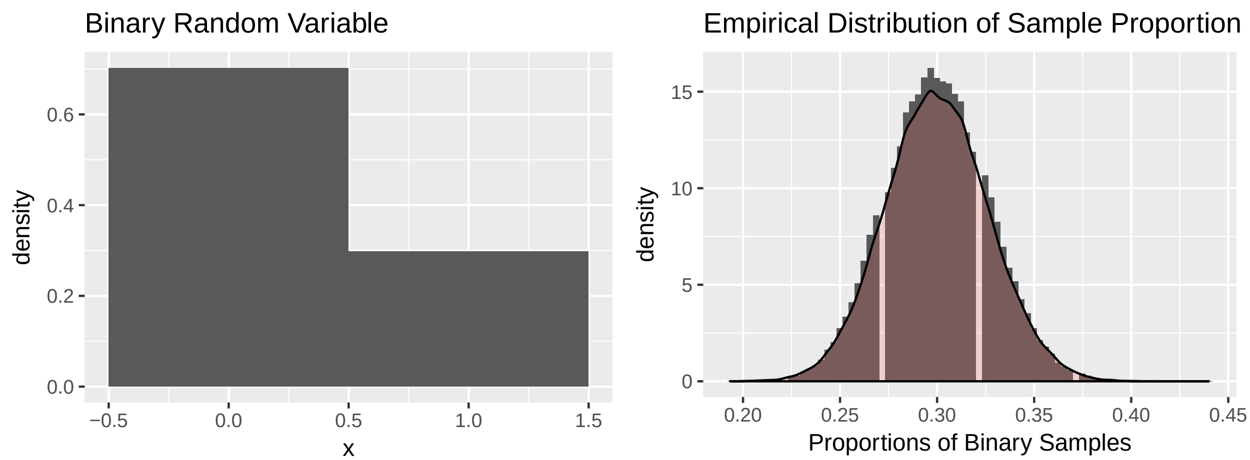
```
## [1] 0.4
```

```
prop (smpl.size = 20, prob = .3, print.smpl = TRUE)
```

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

```
## [1] 0.15
```

```
set.seed(2)
emp.prop <- replicate(n = n.rep, prop(smpl.size = 300, prob = .3), simplify = TRUE)

p8 <- ggplot(mapping = aes(emp.prop) ) +
  geom_histogram(aes(y=..density..), bins = 80) +
  geom_density(alpha = .2, fill="#FF6666") +
  labs(title = "Empirical Distribution of Sample Proportion (n = 300)",
       x = 'Proportions of Binary Samples')

grid.arrange(p7, p8, ncol = 2)
```



Example 2: Sample Mean of a Uniform Sample

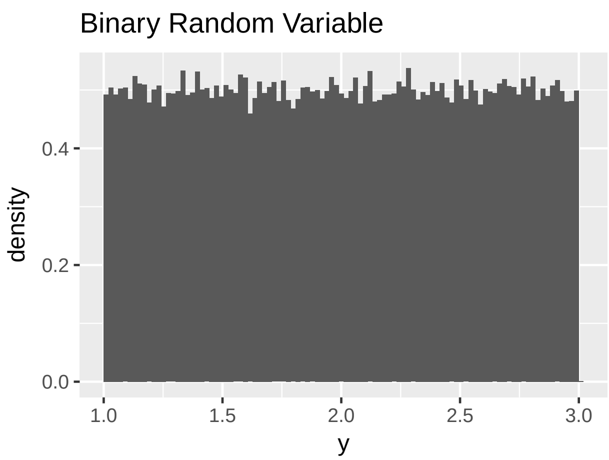
$Y \sim U(1, 3)$

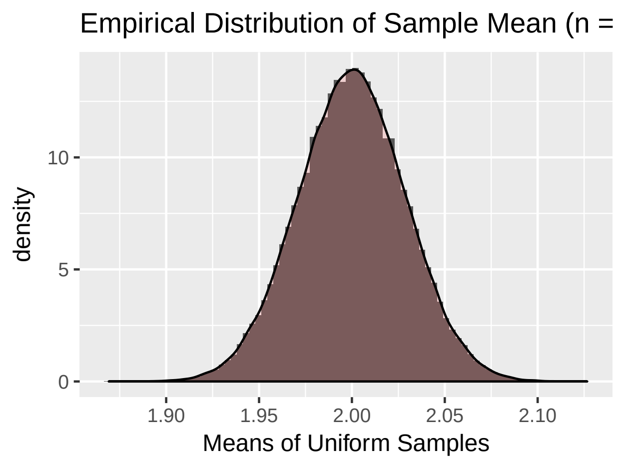
```
set.seed(4)
y <- runif(n.rep, min = 1, max = 3)
p9 <- ggplot(mapping = aes(y) ) + geom_histogram(aes(y=..density..), bins = 100) +
        labs(title = "Binary Random Variable")

unif.mean <- function(smpl.size, min, max, print.smpl = FALSE) {
  smpl <- runif(n = smpl.size, min, max)
  if (print.smpl) print(smpl)
  mean(smpl) %>% return(.)
}

set.seed(666)
emp.mean <- replicate(n = n.rep, unif.mean(smpl.size = 400, min = 1, max = 3), simplify = TRUE)

unif.mean(smpl.size = 20, min = 1, max = 3, print.smpl = TRUE)
```

```
##  [1] 2.675261 1.869563 2.443565 1.945633 1.844290 1.228942 1.733568
##  [8] 2.683585 2.378829 2.563265 1.915734 2.891598 2.215845 2.049039
## [15] 1.858881 1.126504 2.616170 2.459806 1.976874 1.917566
```

```
## [1] 2.119726
```

```
p10 <- ggplot(mapping = aes(emp.mean) ) +
  geom_histogram(aes(y=..density..), bins = 80) +
  geom_density(alpha = .2, fill="#FF6666") +
  labs(title = "Empirical Distribution of Sample Mean (n = 400)",
       x = 'Means of Uniform Samples')

grid.arrange(p9, p10, ncol = 2)
```


Binary Random Variable
density
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
y


Empirical Distribution of Sample Mean (n =
density
10
5
0
1.90
1.95
2.00
2.05
2.10
Means of Uniform Samples