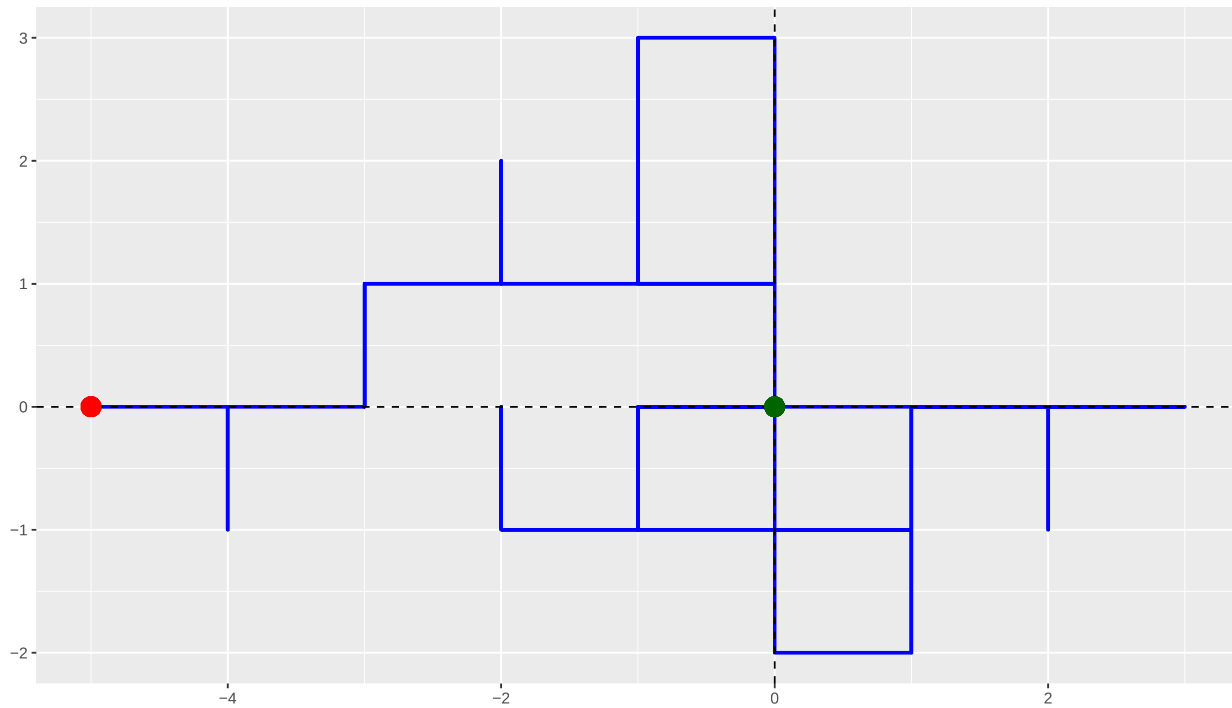
STAT 457 Homework 02

Martha Eichlersmith

2019-10-07

Problem 0.

Graph a random walk



Problem 1a.

Simulate 10,000 random walks of length 50 in $\mathbb{R}^2$ starting at (0, 0). Compute the proportion of walks which end up in each of the four quadrants. What values would you expect for these proportions? Do the observed proportion vary significantly from the expected values?

```
## [1] 0.2479 0.2551 0.2417 0.2553
```

```
## Warning in chisq.test(quad.pct): Chi-squared approximation may be incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  quad.pct
## X-squared = 0.0005096, df = 3, p-value = 1
```

You would expect each quadrant to have approximately 25% of end points for the random walks. The observed proportions do not vary significantly from the expected value.

Problem 1b.

Repeat 1a using walks of length 500

```
## [1] 0.2491 0.2515 0.2454 0.2540

## Warning in chisq.test(quad.pct): Chi-squared approximation may be incorrect

##
##  Chi-squared test for given probabilities
##
## data:  quad.pct
## X-squared = 0.00016088, df = 3, p-value = 1
```

Similar to 1a, the percetnages are very close to 0.25.

Problem 1c.

Repeat 1a and 1b using random walks on a lattice in $\mathbb{R}^2$.

```
## [1] 0.2922 0.2122 0.2851 0.2105

## Warning in chisq.test(quad.pct): Chi-squared approximation may be incorrect

##
##  Chi-squared test for given probabilities
##
## data:  quad.pct
## X-squared = 0.024008, df = 3, p-value = 0.999

## [1] 0.2698 0.2352 0.2646 0.2304

## Warning in chisq.test(quad.pct): Chi-squared approximation may be incorrect

##
##  Chi-squared test for given probabilities
##
## data:  quad.pct
## X-squared = 0.0048336, df = 3, p-value = 0.9999
```

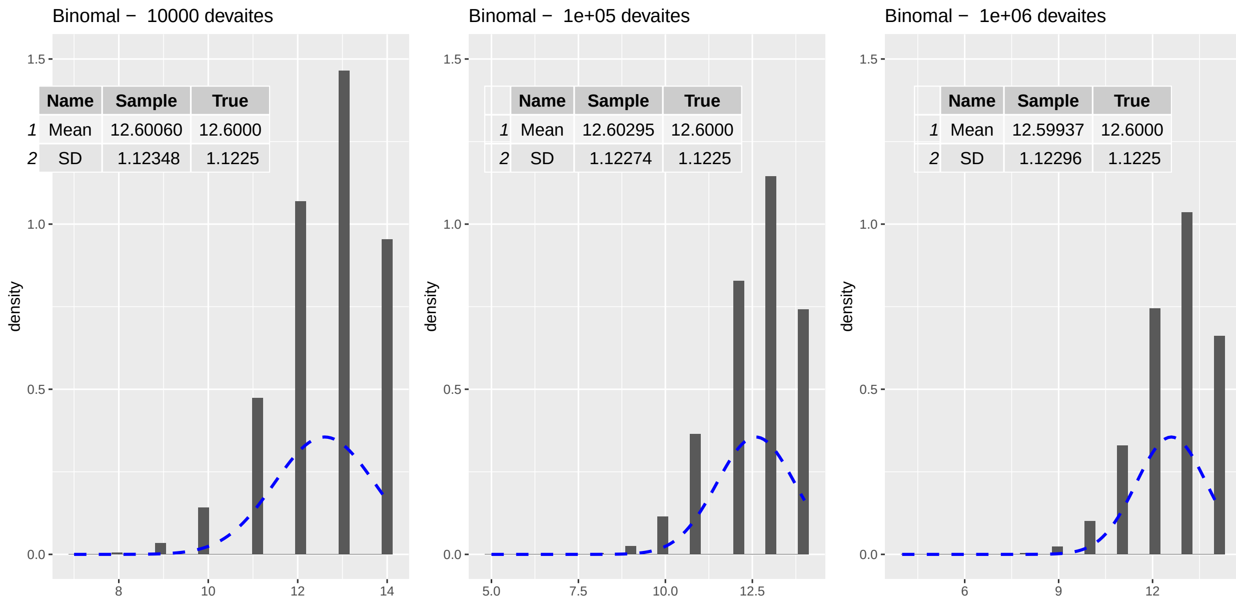
Both of these situations have percentages that are close to 0.25. It is interesting to point out that the χ^2 values are larger for the lattice random walks rather than the continous random walks.

Problem 2.

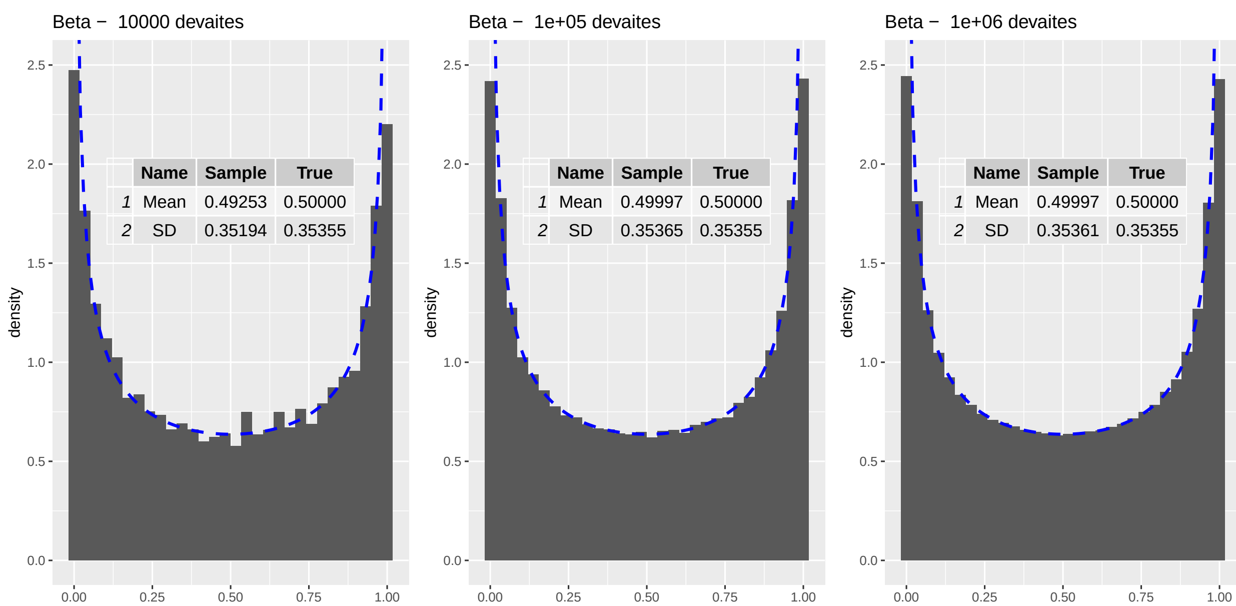
From each of the following distributions: (i) draw 10,000 deviates; (ii) compute the sample average and sample SD and compare these numbers to the true values; (iii) draw a histogram of the deviates along with the plot of the true density.

Binomial(14, .90); Beta(.5, .5); Gamma(12, 2); Inverse.Gamma(12, 2); χ^2 on 2 df; χ^{-2} on 2 df.

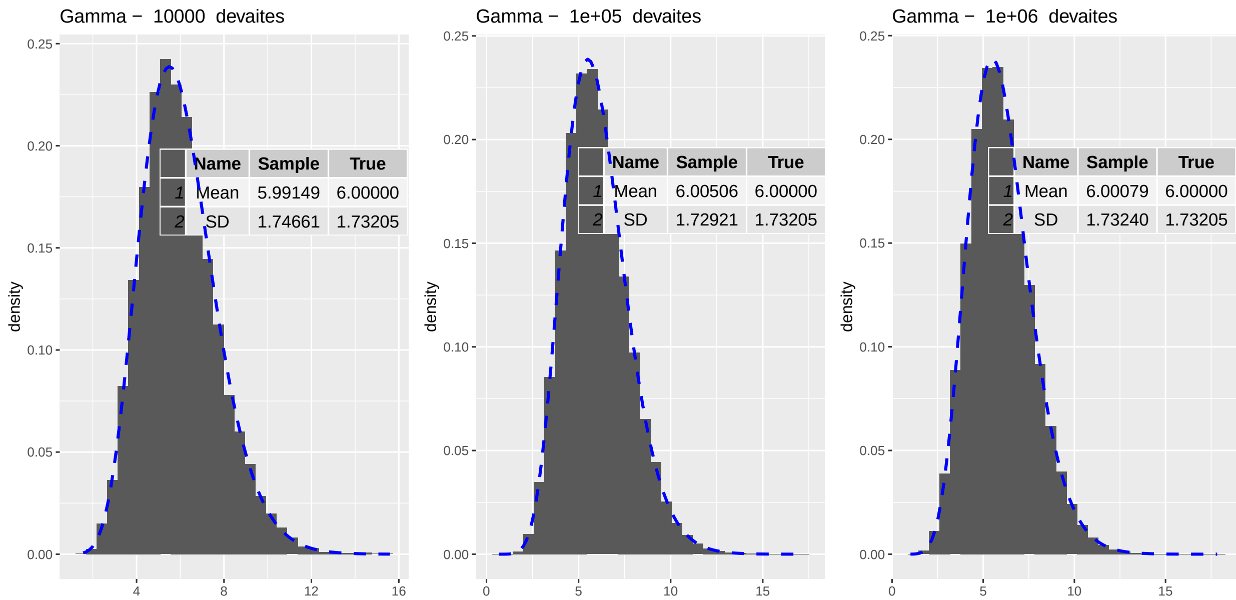
Problem 2a: Binomial(14, .9)



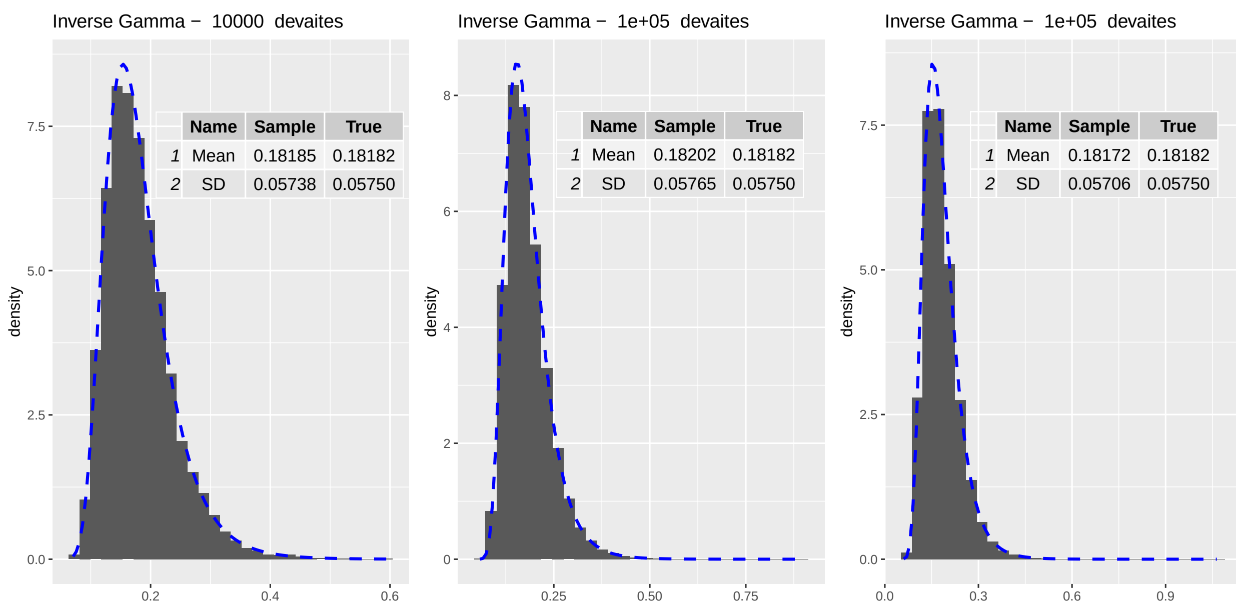
Problem 2b: Beta(.5, .5)



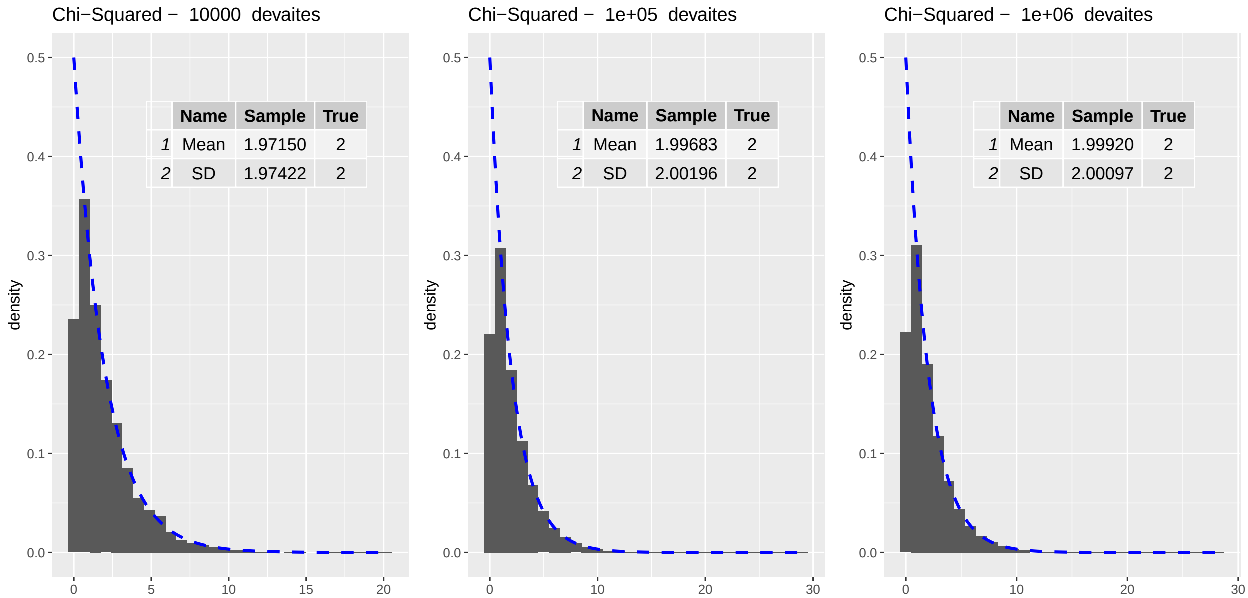
Problem 2c: Gamma(12, 2)



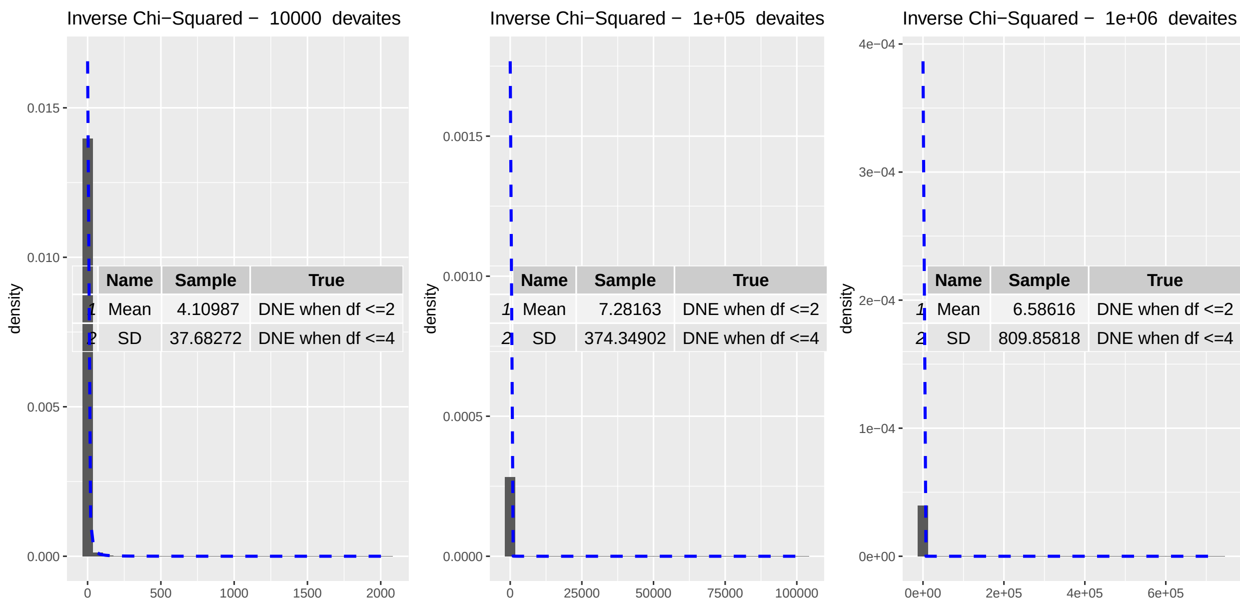
Problem 2d: Inverse.Gamma(12, 2)



Problem 2e: χ^2 on 2 df



Problem 2f: χ^{-2} on 2 df



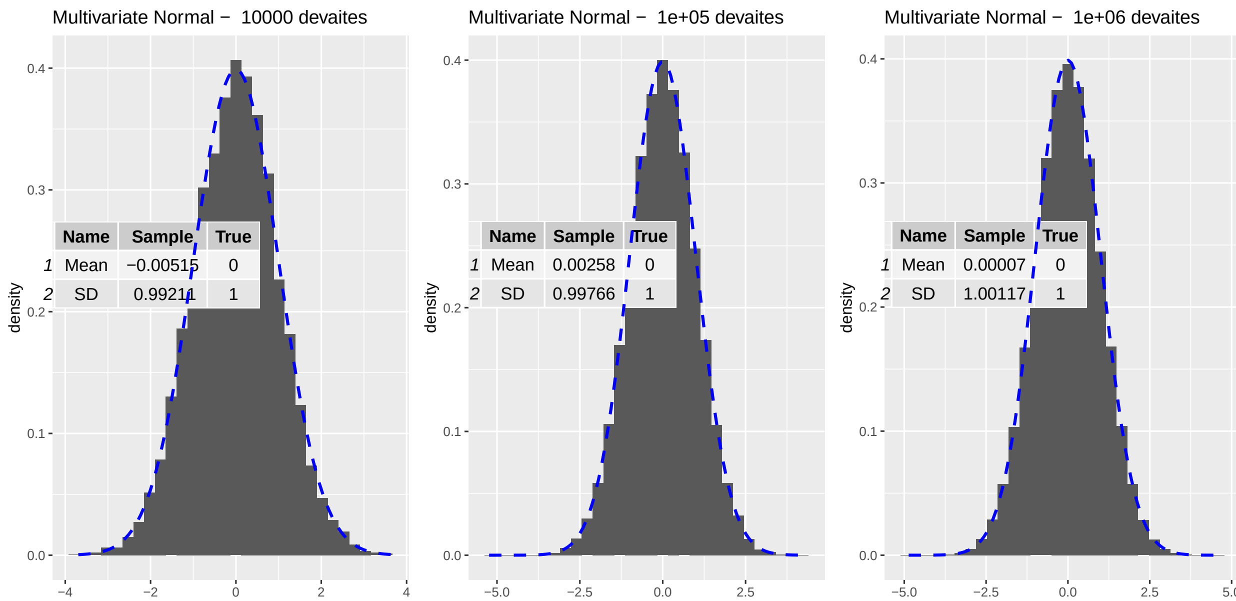
Problem 3a.

Draw a sample size of 10,000 from the trivariate normal distribution with mean (0, 0, 0) and variance-covariance matrix:

$$\begin{bmatrix} 1 & 4.5 & 9 \\ 4.5 & 25 & 49 \\ 9 & 49 & 100 \end{bmatrix}$$

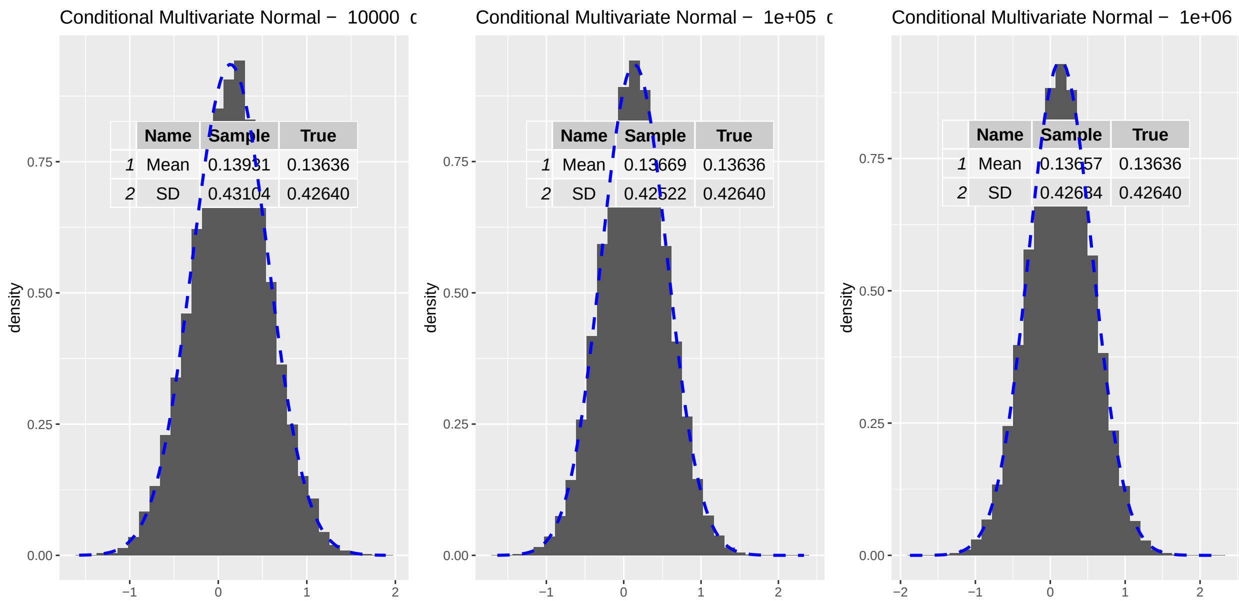
Problem 3b.

Draw a histogram of the X_1 deviates, along with the true marginal. Compute the sample average and sample SD and compare these numbers to the true values.



Problem 3c.

Draw a sample size of 10,000 from the conditional distribution $p(X_1 \mid X_2, X_3)$, take $X_2 = X_3 = 1$. Compute the sample mean and sample SD and compare these numbers to the true values. Draw a histogram of the simulate values, along with the true conditional distribution.



We want to find expected value and sd of $X_1 | X_2 = X_3 = 1$. Let $\boldsymbol{X}_g = (X_2, X_3)$

$$X_1 | \boldsymbol{X}_g \sim N(\mu_1 + \Sigma_{1g}\Sigma_{gg}^{-1}(\boldsymbol{x}_g - \boldsymbol{\mu}_g), \Sigma_{11} - \Sigma_{22}^{-1}\Sigma_{12}^T)$$

$$\begin{aligned}
\mathbb{E}\left[X_1 \mid \boldsymbol{X}_g\right] &= \mu_1 + \Sigma_{1g}\Sigma_{gg}^{-1}(\boldsymbol{x}_g - \boldsymbol{\mu}_g) \\
&= 0 + \begin{bmatrix} 4.5 & 9 \end{bmatrix} \cdot \begin{bmatrix} 25 & 49 \\ 49 & 100 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) \\
&= 0.1363636
\end{aligned}$$

$$\text{s.d.}(X_1 \mid \boldsymbol{X}_g) = 0.4264014$$

Problem 3d.

Use the sample size of 10,000 from the joint trivariate distribution to obtain he conditional mean and SD of X_1 given $X_2 = X_3 = 1$, i.e. look at the triples whose second and third components are within ϵ of 1. How do these values compare to the true mean and SD for various values of ϵ?

```
## [1] "Epsilon of  0.1 : mean =  -0.141384794063079  and sd =  0.559301494177186"
## [2] "Epsilon of  0.5 : mean =  0.145711615569143  and sd =  0.356616710346006"
## [3] "Epsilon of  0.25 : mean =  0.0809213465361516  and sd =  0.359134335696434"
```