ESS 575: Probability Lab 3 - Marginal Distributions

Team England

13 September, 2022

Team England:

- Caroline Blommel
- Carolyn Coyle
- Bryn Crosby
- George Woolsey

cblommel@mail.colostate.edu, carolynm@mail.colostate.edu, brcrosby@rams.colostate.edu, george.woolsey@colostate.edu

Question 1

Fill in Table 2 to estimate the marginal probabilities of presence and absence of the two species. The cells show the joint probability of the events specified in the row and column. The right column and the bottom row show the marginal probabilities.

Table 2: Estimates of marginal probabilities for island occupancy

Events	S	S^c	Marginal
R	$\Pr\left(S,R\right)=\frac{2}{32}$	$\Pr\left(S^c,R\right)=\frac{9}{32}$	$\Pr\left(R\right)=\frac{11}{32}$
R^c	$\Pr\left(S,R^c\right)=\frac{18}{32}$	$\Pr\left(S^c,R^c\right)=\frac{3}{32}$	$\Pr\left(R^c\right)=\frac{21}{32}$
Marginal	$\Pr\left(S\right)=\frac{20}{32}$	$Pr(S^c)=\frac{12}{32}$	$\sum=\frac{32}{32}$

What is the sum of the marginal rows?

The sum of the marginal rows is 1

What is the sum of the marginal columns?

The sum of the marginal columns is 1

Why? Note, when we marginalize over R we are effectively eliminating S and vice versa.

When we marginalize over R we are effecitvely eliminating S because we are interested in making inference on a single parameter, R.

Question 2

Use the data in Table 1 and the probabilities in Table 2 to illustrate the rule for the union of two events, the probability that an island contains either species, $\Pr(R \cup S)$. You will need to derive the formula for the probability of event A or B to solve this problem. A Venn diagram might help you do so.

$$\Pr(R \cup S) = \Pr(R) + \Pr(S) - \Pr(S, R)$$

where:

- $\Pr(R) = \frac{11}{32}$
- $\Pr(S) = \frac{20}{32}$
- $\Pr(S, R) = \frac{2}{32}$

$$\Pr(R \cup S) = \frac{11}{32} + \frac{20}{32} - \frac{2}{32} = \frac{29}{32} = 90.6\%$$

Question 3

Use the marginal probabilities in Table 2 to calculate the probability that an island contains both species i.e., $\Pr(R, S)$, assuming that R and S are independent. Compare the results from those calculations with the probability that both species occur on an island calculated directly from the data in Table 1. Interpret the results ecologically. What is $\Pr(R \mid S)$ and $\Pr(S \mid R)$?

Assuming that R and S are independent

Calculate the probability that an island contains both species: $\Pr(R, S)$

$$\Pr(R, S) = \Pr(R) \cdot \Pr(S)$$

where:

- $\Pr(R) = \frac{11}{32}$
- $\Pr(S) = \frac{20}{32}$

$$\Pr(R, S) = \frac{11}{32} \cdot \frac{20}{32} = 0.215 = 21.5\%$$

What is $\Pr(R \mid S)$ and $\Pr(S \mid R)$?

Assuming that R and S are independent:

$$\Pr(R \mid S) = \Pr(R) = \frac{11}{32}$$

$$\Pr(S \mid R) = \Pr(S) = \frac{20}{32}$$

Assuming that R and S are not independent

Calculate the probability that an island contains both species: $\Pr(R, S)$

$$\Pr(R, S) = \frac{2}{32} = 0.0625 = 6.25\%$$

What is $\Pr(R \mid S)$ and $\Pr(S \mid R)$?

Assuming that R and S are not independent:

$$\Pr(R \mid S) = \frac{\Pr(R, S)}{\Pr(S)} = \frac{\frac{2}{32}}{\frac{20}{32}} = 0.1 = 10\%$$

$$\Pr(S \mid R) = \frac{\Pr(S, R)}{\Pr(R)} = \frac{\frac{2}{32}}{\frac{11}{32}} = 0.182 = 18.2\%$$

Interpretation

Interpret the results ecologically

Assuming that R and S are not independent means that we believe that the presenence of one species influences the presence of the other species and vice versa. The probability that an island contains both species, $\Pr(R, S) = 6.25\%$, means that there is a low chance of both species occupying the same site. Based on this low probability of co-habitation, it is likely that these two species compete with eachother for resources resulting in competitve exclusion.

Continuous random variables

We now explore marginal distributions for continuous random variables. This requires introducing a new distribution, the multivariate normal:

$$\mathbf{z} \sim \mathsf{mulitvariate\ normal}(\mu, \Sigma)$$

where $\mathbf{z}_i$ is a vector of random variables, μ is a vector of means (which can be the output of a deterministic model) and $Sigma$ is a variance covariance matrix. The diagonal of Σ contains the variances and the off diagonal contains the covariance of $\Sigma[i, j]$. The covariance can be calculated as $\sigma_i \sigma_j \rho$ where σ_i is the standard deviation of the ith random variable, σ_j is the standard deviation of the jth random variable, and ρ is the correlation between the random variable i and j. The covariance matrix is square and symmetric. We will learn more about these matrices later in the course. For now, an example will go a long way toward helping you understand the multivariate normal distribution.

The rate of inflation and the rate of return on investments are know to be positively correlated. Assume that the mean rate of inflation is `.03` with a standard deviation of `0.015`. Assume that the mean rate of return is `0.0531` with a standard deviation of `0.0746`. Assume the correlation between inflation and rate of return is `0.5`.

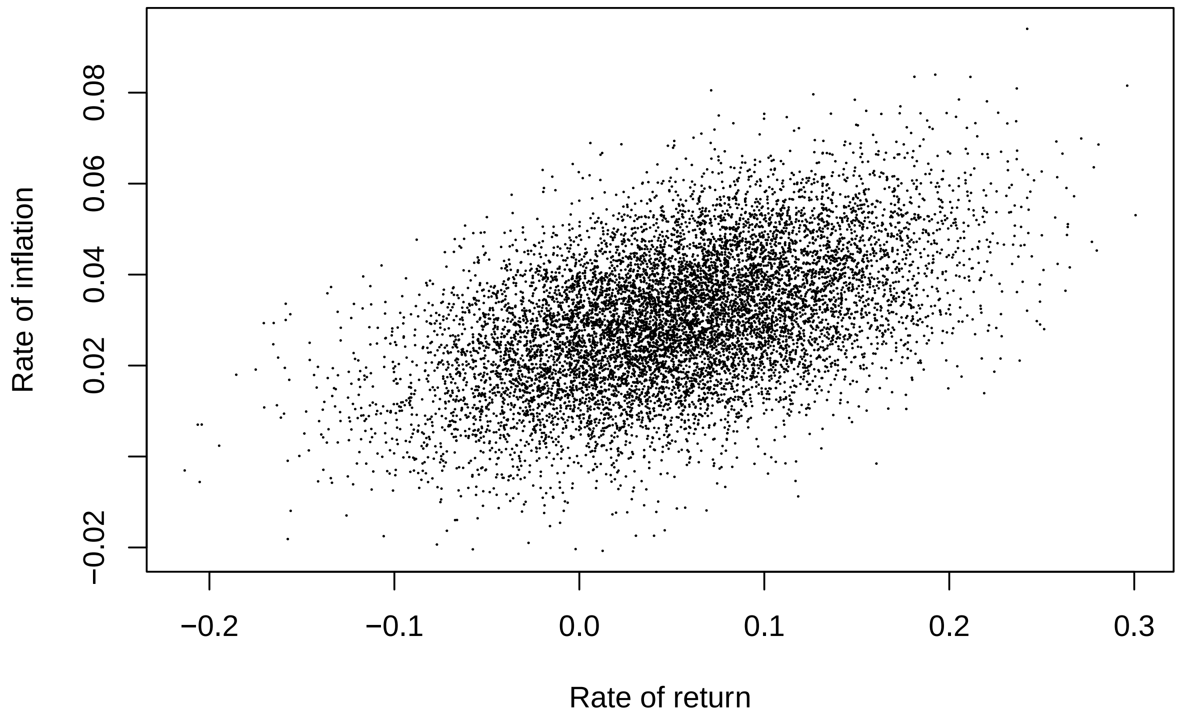
You can simulate interest rate and inflation data reflecting their correlation using the following function:

```
DrawRates = function(n, int,int.sd, inf, inf.sd, rho.rates) {
  covar = rho.rates * int.sd * inf.sd
  Sigma <- matrix(c(int.sd^2, covar, covar, inf.sd^2), 2, 2)
  mu = c(int,inf)
  x = (MASS::mvrnorm(n = n, mu = mu, Sigma))
  return(x)
}

mu.int = .0531
sd.int = .07 #.0746
mu.inf = .03
sd.inf = .015 #.015
rho=.5
```

```r
n = 10000

x = DrawRates(n = n, int = mu.int, int.sd = sd.int, inf = mu.inf, inf.sd = sd.inf, rho.rates = rho)
par(mfrow=c(1,1))
plot(x[, 1], x[, 2], pch = 19, cex = .05, xlab = "Rate of return", ylab = "Rate of inflation")
```

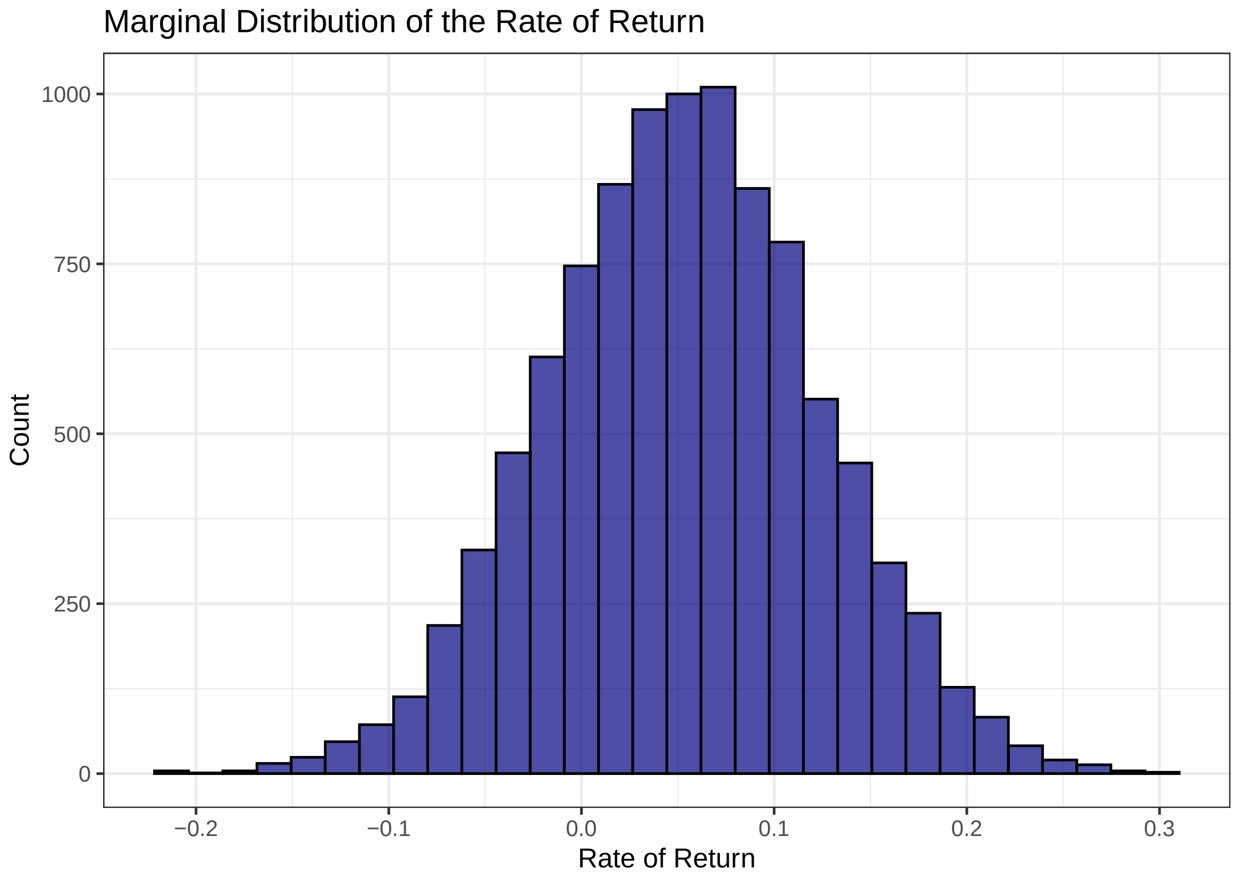


What would this cloud look like if the rates were not correlated?

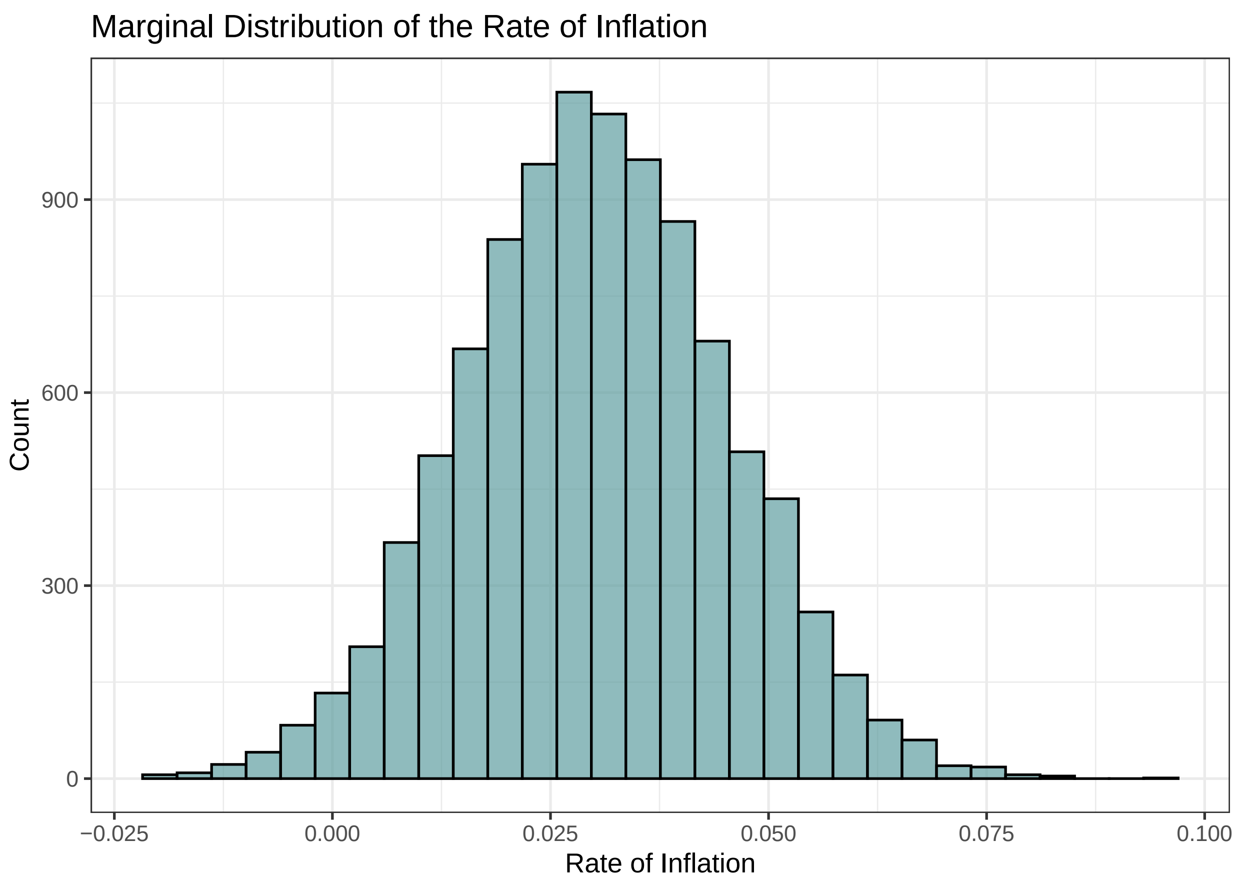
If the rates were not correlated then the data would be dispersed randomly in the plot.

Show an approximate plot of the marginal distribution of each random variable. It turns out this is the way we will do it with MCMC.

```r
# create data frame
dta <- data.frame(
  return_rate = x[,1]
  , inflation_rate = x[,2]
)
# return rate distribution
ggplot(data = dta) +
  geom_histogram(aes(return_rate), fill = "navy", lwd = 0.5, color = "black", alpha = 0.7) +
  xlab("Rate of Return") +
  ylab("Count") +
  labs(title = "Marginal Distribution of the Rate of Return") +
  theme_bw()
```



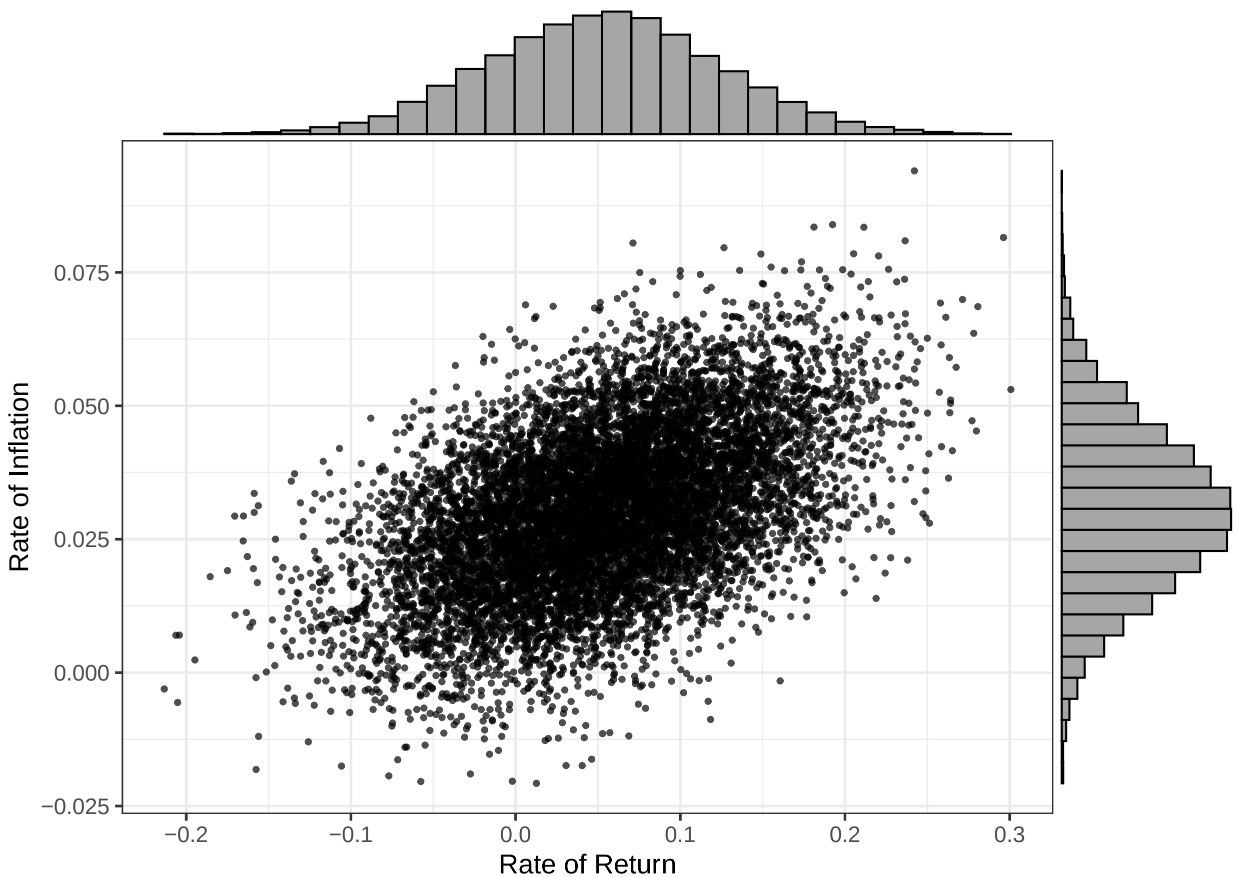

```r
# inflation rate distribution
ggplot(data = dta) +
  geom_histogram(aes(inflation_rate), fill = "cadetblue", lwd = 0.5, color = "black", alpha = 0.7) +
  xlab("Rate of Inflation") +
  ylab("Count") +
  labs(title = "Marginal Distribution of the Rate of Inflation") +
  theme_bw()
```



Can you combine the histogram with the scatter plot to more clearly reveal the marginalization? Take a look at the `ggExtra` package and `ggMarginal` function.

```
p1 <- ggplot(dta, aes(x = return_rate, y = inflation_rate)) +
  geom_point(alpha = 0.7, size = 0.7) +
  xlab("Rate of Return") +
  ylab("Rate of Inflation") +
  theme_bw()

p1 %>%
  ggExtra::ggMarginal(type = "histogram", fill = "gray50", alpha = 0.7, lwd = 0.4)
```


Rate of Inflation
0.075
0.050
0.025
0.000
−0.025
−0.2
−0.1
0.0
0.1
0.2
0.3
Rate of Return