

Finding the expected value of order statistics can be a challenge if one is not privy to certain mathematical “tricks” and techniques. Thankfully, by use of simulation we can have some guidance that can help us realize when we are wrong with our mathematical conclusions and when we are correct.

This post was inspired by an assignment question I had in my mathematical statistics class and how we were able to catch our miscalculations by simulating the distribution. I am very grateful to my colleagues who showed me this technique – I see that this will be useful in future applications!

The Question

$(X_1, X_2)^T$ has a bivariate normal distribution, with $E(X_1) = E(X_2) = 0$, and $Var(X_1) = Var(X_2) = 1$, and $Cov(X_1, X_2) = \rho$. Find $E(\max\{X_1, X_2\})$.

Looking back at this question now, it looks quite trivial. But when we were trying to solve it- we were consistently (and wrongly) concluding that $E(\max\{X_1, X_2\}) = 0$ until we turned to a tried and true tool for studying distributions- *simulation*

Using the {mvtnorm} package

Thankfully, the {mvtnorm} package has been around for awhile and specializes in simulating multivariate normal distributions. With this package, we can get away with simulating the bivariate normal distribution and examine its relationship with the correlation coefficient, ρ , as it increases.

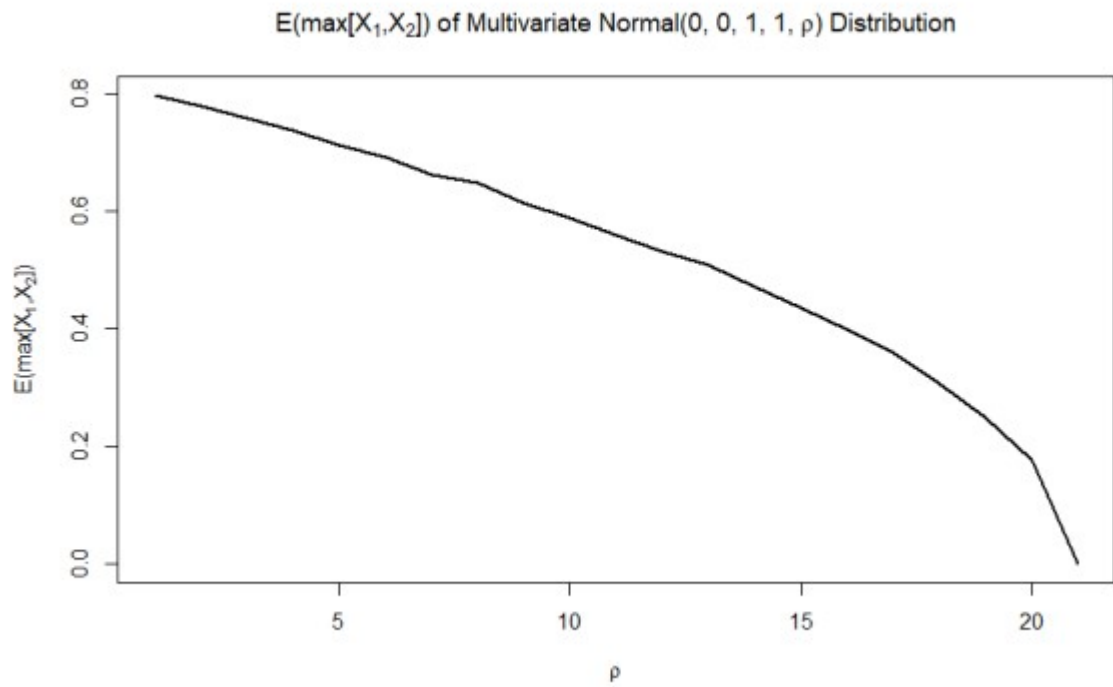
```
# Simulating the Expected value of E(max{X,Y}) where (X,Y)~ Bivariate Normal
Density
# as the Correlation Coefficient increases
library(mvtnorm)
set.seed(1234)

# #The Correlation Coefficient test values

rho <- seq(-1, 1, by=0.1)
# Sample size

n <- 100000
resMax <- sapply(rho, function(t) {
  x <-
    rmvnorm(n, sigma = matrix(
      data = c(1, t, t, 1),
      nrow = 2,
      ncol = 2,
      byrow = TRUE
    ))
  z <- mapply(function(u, v) {
    max(u, v)
  }, u = x[, 1], v = x[, 2])
  mean(z)
})
```

Plotting the data reveals that our expected value takes a range of values depending on ρ !

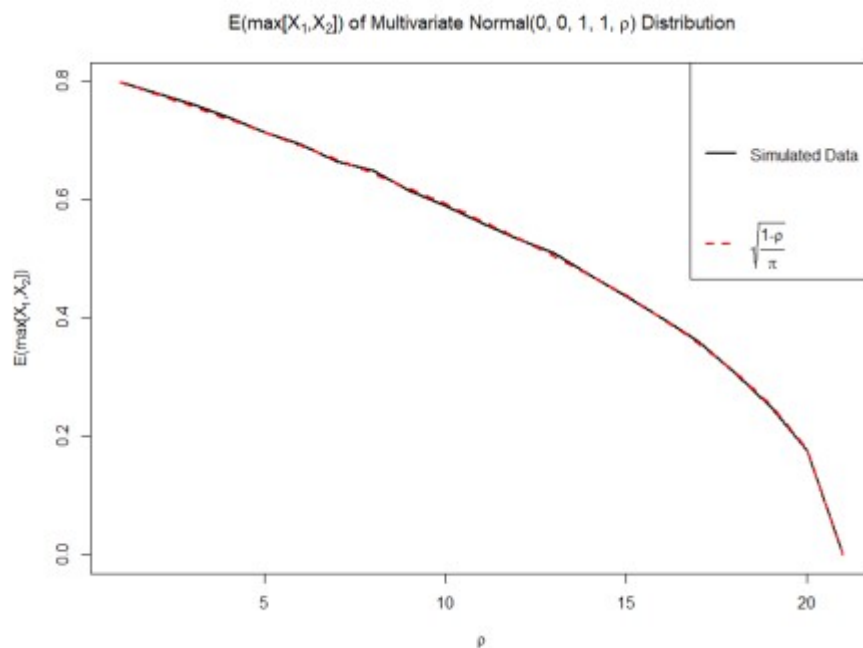


With this in mind, we were able to determine by reevaluating the problem that

$$E(\max\{X_1, X_2\}) = \sqrt{\frac{1-\rho}{\pi}}$$

We are able to confirm this by plotting the true expected value against our simulated data.

```
trueVals <- sapply(rho, function(t) {
  x <- sqrt((1 - t) / pi)
})
```



And there you have it!

Conclusion

When working through symbolic calculations one can be prone to running into errors. Simulation is a valuable tool in this setting as it helps provide insight to where the true answer is- especially when the underlying distribution is known.