SIMULATION IN FINANCE

1

FRE 6103

Lecture 4 / Chapters 5-6

Professor David Shimko

TODAY'S TOOLKIT

- Simulating any random variable
- Estimating the mean of a function of random variables

or stdev or probability distribution e.g. $f(z) = \exp(z)$ with $E[f(z)] = \exp(0.5)$

- Error ranges

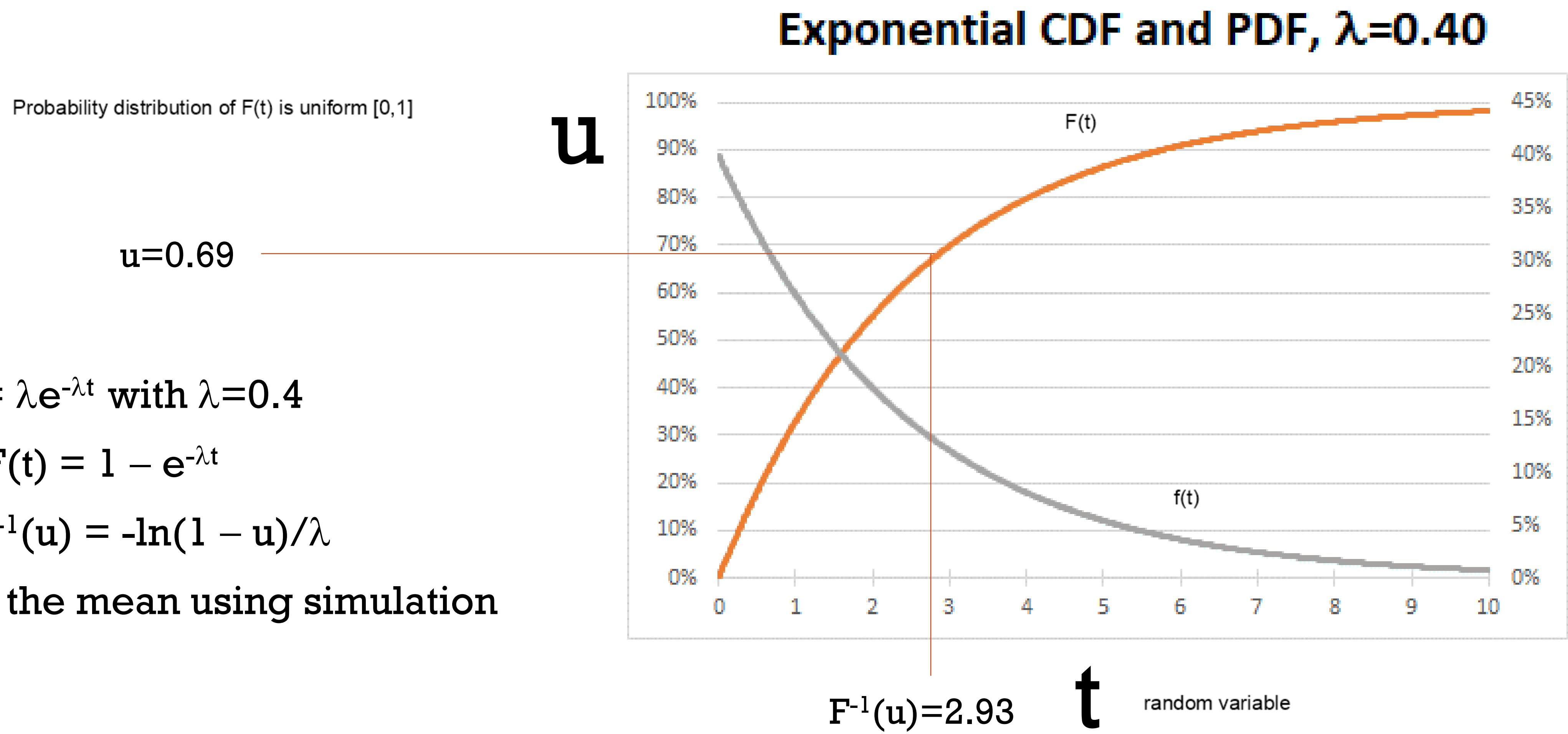
ALWAYS INCLUDE!

ex. Indicative deal price vs firm price (prepared to contract)

- Variance reduction
- Simulating correlated random variables
- Nonlinear "correlations"

Miniproject 3: Correlated exponential random variables

EXPONENTIALLY DISTR RANDOM NUMBERS



- $f(t) = \lambda e^{-\lambda t}$ with $\lambda=0.4$
- $u = F(t) = 1 - e^{-\lambda t}$
- $t = F^{-1}(u) = -\ln(1 - u)/\lambda$
- Find the mean using simulation

NORMALLY DISTR RANDOM NUMBERS

*used to generate correlated random numbers, even if not normally distributed

- $z = N^{-1}(u)$
- z = normsinv(rand( ))

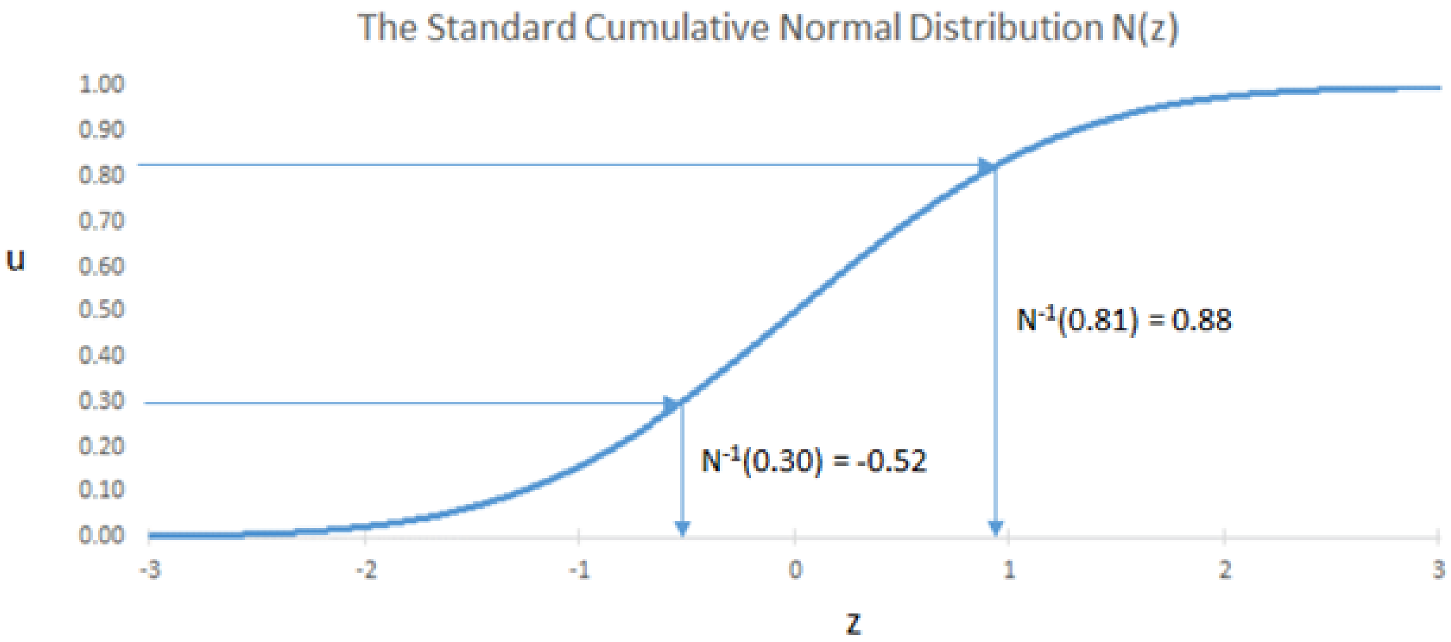
  standard

- Can you "correct" the sample by making the mean 0 and standard deviation 1?

  Subtract mean of sample, divide by stdev of the sample

- What does this accomplish?

  1. Eliminate one source of uncertainty from the simulation
  2. Good for client presentations -- align your mean results to client's assumptions



F-DISTRIBUTED RANDOM NUMBERS (10,6)

=FINV(rand(),10,6)

- Simulate N=1000 values
- What is the mean estimate from your sample?
- What is the standard deviation estimate from your sample?
- What is the "maximum error" in your mean estimate?
- How does this compare with the true mean of the distribution?
- Why might your range not include the true mean?

EVERY SIMULATION DOES IT LIKE THIS

- Generate a large number N of random numbers or vectors z
- Compute outcome values f(z)

function of random numbers

- Compute the average value $\bar{X}$ of the f(z) results --- this is your estimate of E[f(z)]
- Compute the standard deviation of the values of f(z). Let's call this "s"
- Compute the *standard error*, which is the standard deviation divided by $\sqrt{N}$, i.e. $s/\sqrt{N}$
- Compute the maximum likely error, which is m times the standard error $= ms/\sqrt{N}$ (m is a matter of choice corresponding to the number of standard errors)
- If the maximum likely error is too large, increase the size of N in proportion to the square of the desired increase in precision.
- The solution is then given by $E[f(z)] = \bar{X} \pm ms/\sqrt{N}$

max likely error

s^2 = var(X)
but I want variance of X-bar, the sample average
Var xbar = Var (sum of Xs divided by N) = Var(sum of X's)/N^2 = N s^2/ N^2 = s^2/N
Stdev of mean estimate = s/sqrt(N)

Central Limit Theorem - If X has any probability distribution, then the sample X bar aproaches a normal distribution as N grows large

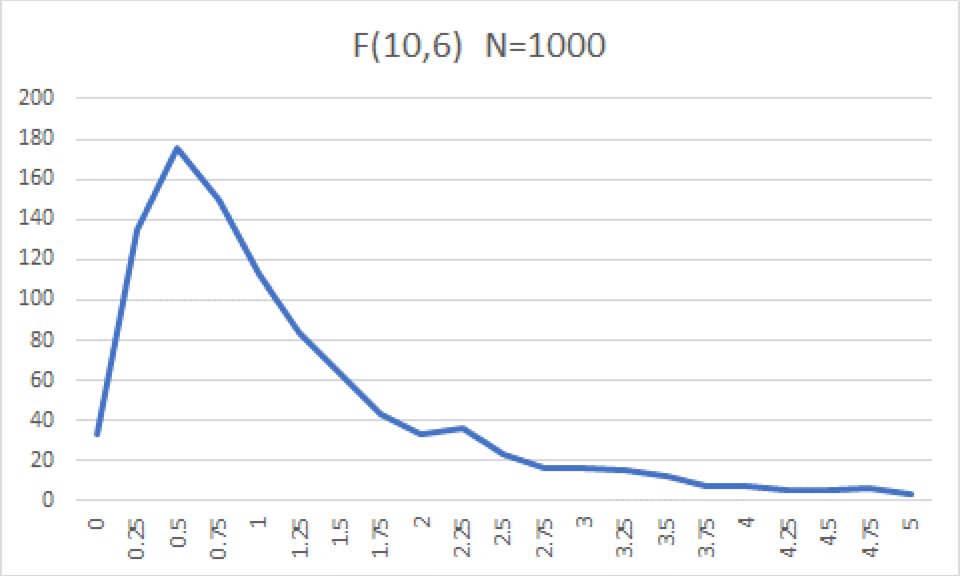
Use m=2 in this class!

GRAPH THE PROBABILITY DISTRIBUTION

- Use your F(10,6) distribution values to construct a frequency distribution
- Recommended parameters
  - 20 bins, interval 0.25 from 0-5

In Excel

ANALYSIS toolpak - static output
COUNTIF function - dynamic output (adjusts to changing inputs)



VARIANCE REDUCTION

- These are techniques designed to increase simulation precision for the same number of sims, or reduce the number of sims required
- Class example: We want to find the value of $E[e^z]$
  - z is the standard normal variate   mean 0, stdev 1
  - Start with 2500 simulations.
    - What is your estimate?
    - What should it be?
    - What is the maximum error in the value?

VARIANCE REDOX: ANTITHETIC

- Can you improve the model with antithetic sampling?

Double sample size by includind
negative complements of z

		RAW	ANTITHETIC
Mean	-0.01	1.60	1.61
Stdev	0.97	2.02	2.00
Count	2500	2500	5000
Maxerr		0.08	→ 0.06
Div sqrt 2		0.06	

	z	ez	e(-z)
1	0.24	1.28	0.78
2	0.54	1.71	0.58
3	0.97	2.65	0.38
4	-0.06	0.95	1.06
5	0.62	1.86	0.54
6	1.53	4.64	0.22
7	1.33	3.79	0.26
8	0.35	1.42	0.71
9	-0.97	0.38	2.63
10	1.12	3.05	0.33

...2500 rows

VARIANCE REDOX: MOMENT MATCHING

- Can we improve the model by matching moments?
- Transform your z values so that the mean is 0 and the standard deviation is 1.
- Does this help?

STDIZE	MOMENT	ANTI+MOMENT
0.00	1.65	1.65
1.00	2.21	2.24
	2500	5000
	0.09	0.06
	ez	e(-z)
-0.20	0.82	1.23
0.78	2.19	0.46
0.15	1.16	0.86
-3.51	0.03	33.40
0.94	2.57	0.39
-0.65	0.52	1.92
0.16	1.17	0.86
0.58	1.79	0.56
0.48	1.61	0.62
-0.39	0.67	1.48

VARIANCE REDOX: CONTROL VARIATE

- Can we predict $E(e^z)$ better by using z and z^2 as (linear) controls?
  - Write $E[e^z] = b_0 + b_1 E[z] + b_2 E[z^2] + \cancel{\varepsilon}$
  - We know $E[z] = 0$ and $E[z^2] = 1$
  - Simulate z values and perform the regression
  - Estimated result =
    - 0.84 + 0 + 0.80 = 1.64
    - Correct answer = 1.648721 = $e^{0.5}$

b = { b0 b1 b2 }

add constant term	x		y
1	z	z2	ez
1.00	-0.42	0.18	0.66
1.00	1.03	1.06	2.80
1.00	-1.83	3.34	0.16
1.00	0.40	0.16	1.49
1.00	0.80	0.63	2.22
1.00	1.38	1.91	3.98
1.00	1.41	1.98	4.09
1.00	-0.53	0.29	0.59
1.00	1.09	1.20	2.99
1.00	-0.59	0.35	0.55

Regression statistics (updates with every sim)

	LINEST	regress y on x - multivariate	
	b2	b1	b0
Coeff	0.7966	1.4591	0.8408
Sterr	0.0346	0.0415	0.0513
R2/se	0.96	0.40	
ANOVA	1087.05	97.00	
	340.70	15.20	

Older version of Excel - Shade in whole box, type the formula, then Ctl+Shift+Enter

Matrix notation for variance and covariance

Var(aX + bY) = a^2 Var(X) + 2ab cov(X,Y) + b^2 Var(Y)

or in matrix notation, let cov matrix be "cov" (2x2 matrix)

Variance = (a b) cov (a b)'

Cov(aX+bY,cX+dY) = (a b) cov (c d)'

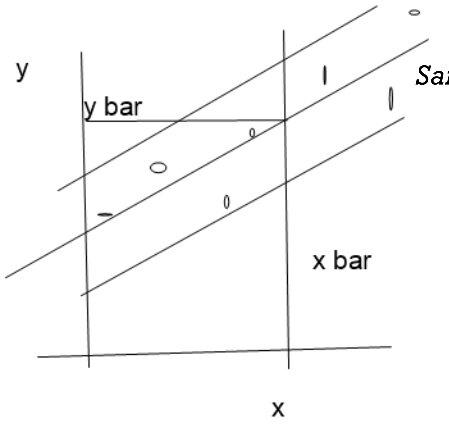
CONTROL VARIATE MAXIMUM ERROR

- Use the "variance of the mean response" to calculate the standard error of the mean estimate in regression
- Define the data matrix X as the pink box to the right
- Let $b' = (b_0 \quad b_1 \quad b_2)$
  - The estimated regression coefficients
- Let $X_h' = (1 \quad 0 \quad 1)$ any value of X
  - X_h could be any of the rows of the X matrix
  - Or any vector X of interest
  - We are interested in using the expected values of the columns
- Then the estimated mean of e^z is $\mathrm{E}[e^z] = X_h' b$.
  - X_h is given but b is measured with error
- The variance of the mean is $Var(\mathrm{E}[e^z]) = s_e^2 X_h' (X'X)^{-1} X_h$

response

  - s_e^2 is the residual variance from the regression
  - The covariance (matrix) of b is $Cov(b) = s_e^2 (X'X)^{-1}$
  - Don't forget to take the square root!

prediction of y = a + bx +/- 2*se



Sample data

1	z	z2	ez
1.00	-0.42	0.18	0.66
1.00	1.03	1.06	2.80
1.00	-1.83	3.34	0.16
1.00	0.40	0.16	1.49
1.00	0.80	0.63	2.22
1.00	1.38	1.91	3.98
1.00	1.41	1.98	4.09
1.00	-0.53	0.29	0.59
1.00	1.09	1.20	2.99
1.00	-0.59	0.35	0.55

if y = a + bx + e, then Var(y|x) = Var(a) + 2X cov(a,b) + X^2 Var(b) + se^2

in simple regression, Var(b) = se^2/Var(x)

CALCULATION ILLUSTRATION

- Take N=100. We know the true value of $X'X$ and therefore its inverse
- We also know the true value of $X'Y$ and therefore *b* and *Cov(b)*

true	100	0	100	true	0.015	0.000	-0.005
X'X	0	100	0	X'X inv	0.000	0.010	0.000
	100	0	300		-0.005	0.000	0.005
true	164.87 =EXP(0.5)*100			true	0.0089	0.0000	-0.0030
X'Y	164.87			Cov(b)	0.0000	0.0059	0.0000
	329.74 =2*EXP(0.5)*100				-0.0030	0.0000	0.0030

True b values

X'y	b
164.87	0.8244
164.87	1.6487
329.74	0.8244

Analytical SD of mean response

Xh(X'X)(-1)Xh	0.0100
times se2	0.0059
SD of mean response	0.0770

Computation of true se2

Residual	738.91	SSY
	-1359.14	-2Cov(Y,Y hat)
	679.57	Var(Y hat)
Sum	59.34	Var resid
	0.77	SD resid

MULTIVARIATE SIMS

- Can always simulate multiple uncorrelated variables, the problem is to correlate them
  - *What is the meaning of correlation in this context?*
  - *How do we handle more than a few correlated normal variables?*
  - *How do we handle non-normal variables?*
- Correlation is defined relative to a linear relationship
  - Does not apply to all joint probability distributions
  - Constant parameters (mean, stdev, correl)

ex. u = a + bz would give
outcomes not in the range [0.1]

Pearson Correlation = cov(x,y)/(sx*sy)

Properties:
1. Each x,y pair is independent of every other pair
2. Assume sx, sy, meanx, meany, corr constant
3. x,y are jointly normal
4. E[y|x] = linear function of x

Correlation means something
different outside of normal
distributions.

BIVARIATE NORMAL SIMS

- Suppose we have a standard normal variate z_1.
- Can we create z_2 so that its mean is zero, stdev is 1, and correlation with z_1 is ρ?

  normally distributed
- Take z_3 independent of z_1
- Let $\boxed{z_2 = \rho z_1 + z_3\sqrt{1-\rho^2}}$
- Test the mean, stdev & correl

  E[z2] = 0
  Var(z2) = rho^2 + (1 - rho^2) = 1
  Cov(z1,z2) = Cov(z1,rho*z1 + z3*radical) = rho + 0 = rho
- In matrix form, $\underbrace{(z_1 \quad z_2)}_{\text{correl}} = \underbrace{(z_1 \quad z_3)}_{\text{uncorr}} \begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{pmatrix}$
- **What is this 2x2 matrix called?**

Transpose of
the Cholesky Decomposition

Cholesky definition - lower diagonal

ex. Covariance matrix
Cholesky decomp x Transpose of
cholesky = Covariance matrix
("square root of the cov matrix")

ALSO uncorr = correl x inverse(cholesky transpose)

MULTIVARIATE NORMAL SIMS (5)

- Class project:
  - Create a set of 100 instances of 5 random normal variates that have an *exact* target correlation matrix
  - Target correlation matrix = Common correlation 0.2 for all off-diagonal elements
  - Test your results

	Common correlation				0.2
Target	1.00	0.20	0.20	0.20	0.20
	0.20	1.00	0.20	0.20	0.20
	0.20	0.20	1.00	0.20	0.20
	0.20	0.20	0.20	1.00	0.20
	0.20	0.20	0.20	0.20	1.00
CHOLESKY	1.00	0.00	0.00	0.00	0.00
	0.20	0.98	0.00	0.00	0.00
	0.20	0.16	0.97	0.00	0.00
	0.20	0.16	0.14	0.96	0.00
	0.20	0.16	0.14	0.12	0.95
TEST	1.00	0.20	0.20	0.20	0.20
	0.20	1.00	0.20	0.20	0.20
	0.20	0.20	1.00	0.20	0.20
	0.20	0.20	0.20	1.00	0.20
	0.20	0.20	0.20	0.20	1.00

File: Lecture 4 - Cholesky five.xlsx

COPULAE

- Class exercise:
  - "Correlate" two random uniform variates
  - Use this to correlate two exponential distributions with parameters 0.2 and 0.3
  - Show a scatterplot of your results
    - For zero correlation
    - For 80% correlation

Steps:

1. Simulate uncorrelated normal numbers
2. Force a correlation (moment matching)
3. Convert to uniform (result: "correlated" uniforms)
4. Convert to correlated exponentials using F(-1)(u)
5. Graph

CONCLUSIONS

- Any random variable can be simulated by the function $F^{-1}(u)$, where F^{-1} is the inverse CDF of u, and u is a uniformly distributed random variable on [0,1]
- To simulate the value of a function of number of uncorrelated random variables (any distributions), follow these steps
  - Generate N values and calculate the mean
  - Calculate the standard deviation, and divide by $\sqrt{N}$ to get the standard error
  - Report the mean as your estimate and your "maximum error" as 2 times the standard error
  - Increase the sample size if more precision is needed
- To correlate normally distributed random variables, use independent random simulations and multiply the data vector by the transpose of the Cholesky decomposition of the desired covariance matrix
  - (Pearson) Correlation is only defined for normal random variables
- To "correlate" random variables of any distribution, convert correlated normal random numbers to uniform distributions, and then use $F^{-1}(u)$ to produce the desired random variables