How To Correctly Simulate Multiple Random Variables

Probability and Statistics for Data Science

Carlos Fernandez-Granda

NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

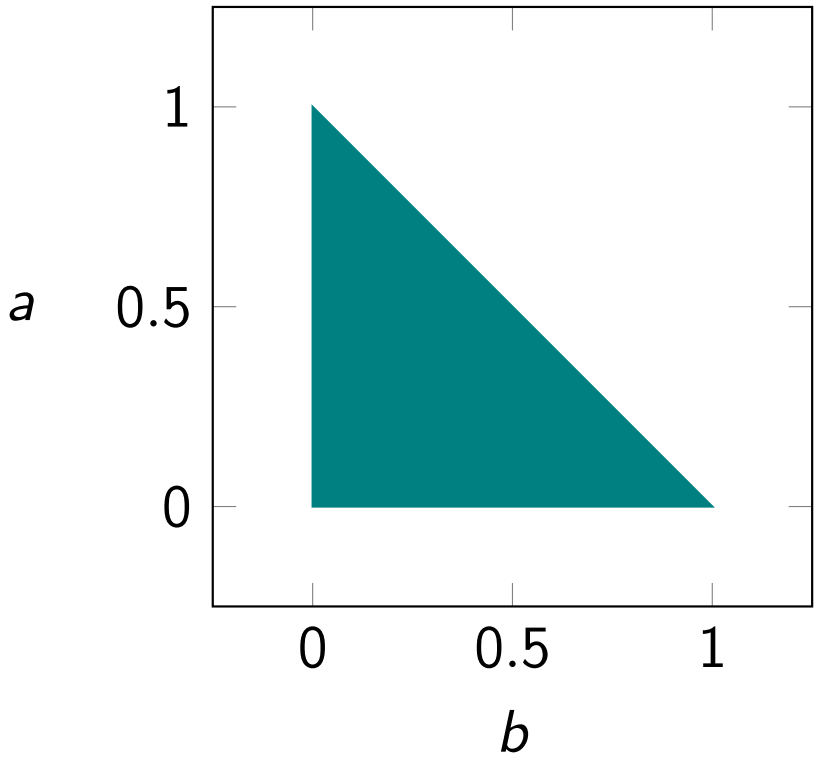
NYU DATA SCIENCE

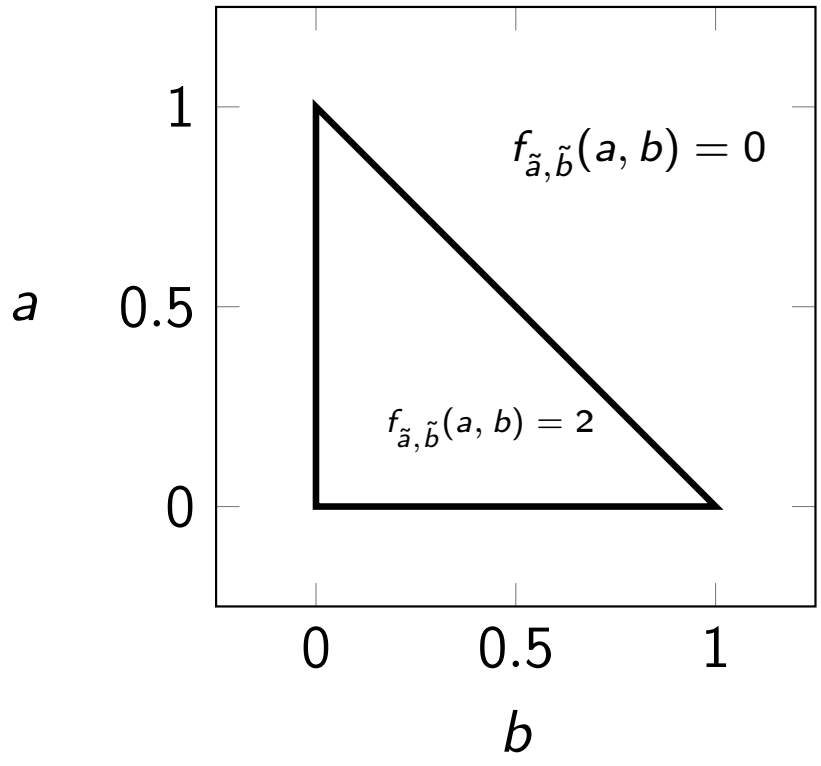
These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Goal

Simulate joint distribution of multiple random variables

Triangle lake: Joint pdf



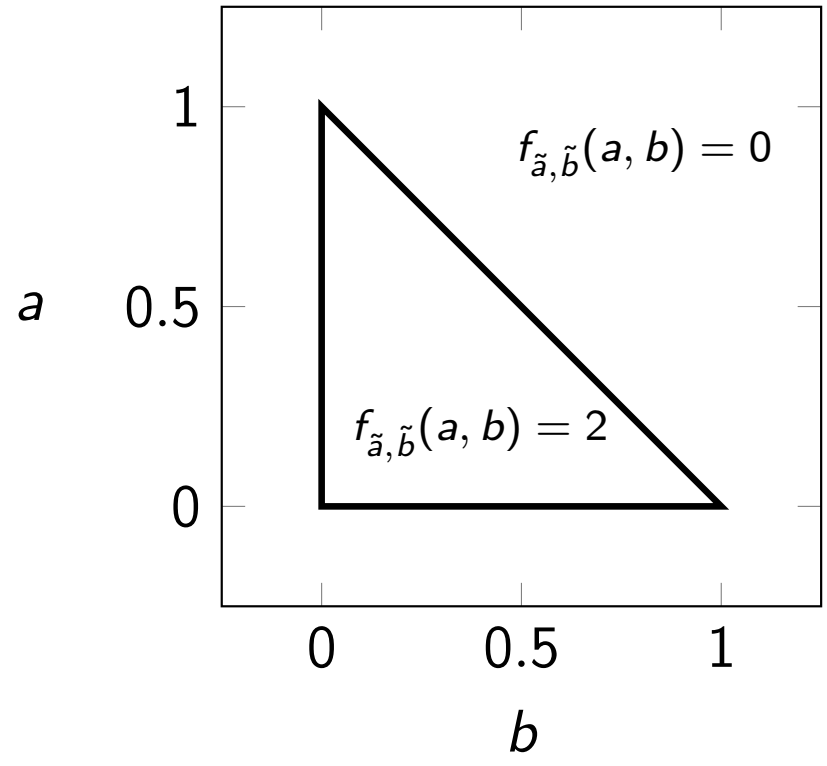


Simulating a single random variable

Inverse-transform sampling: To generate samples from random variable $\tilde{a}$ with cdf $F_{\tilde{a}}$ we

1. Generate sample u from uniform random variable $\tilde{u}$
2. Set $a := F_{\tilde{a}}^{-1}(u)$

Marginal distribution of $\tilde{a}$



$$f_{\tilde{a}}(a) = \int_{b=-\infty}^{\infty} f_{\tilde{a},\tilde{b}}(a,b)\,\mathrm{d}b = \int_{b=0}^{1-a} 2\,\mathrm{d}b = 2(1-a)$$

$$F_{\tilde{a}}(a) = 2a - a^2 \qquad F_{\tilde{a}}^{-1}(u) = 1 - \sqrt{1-u}$$

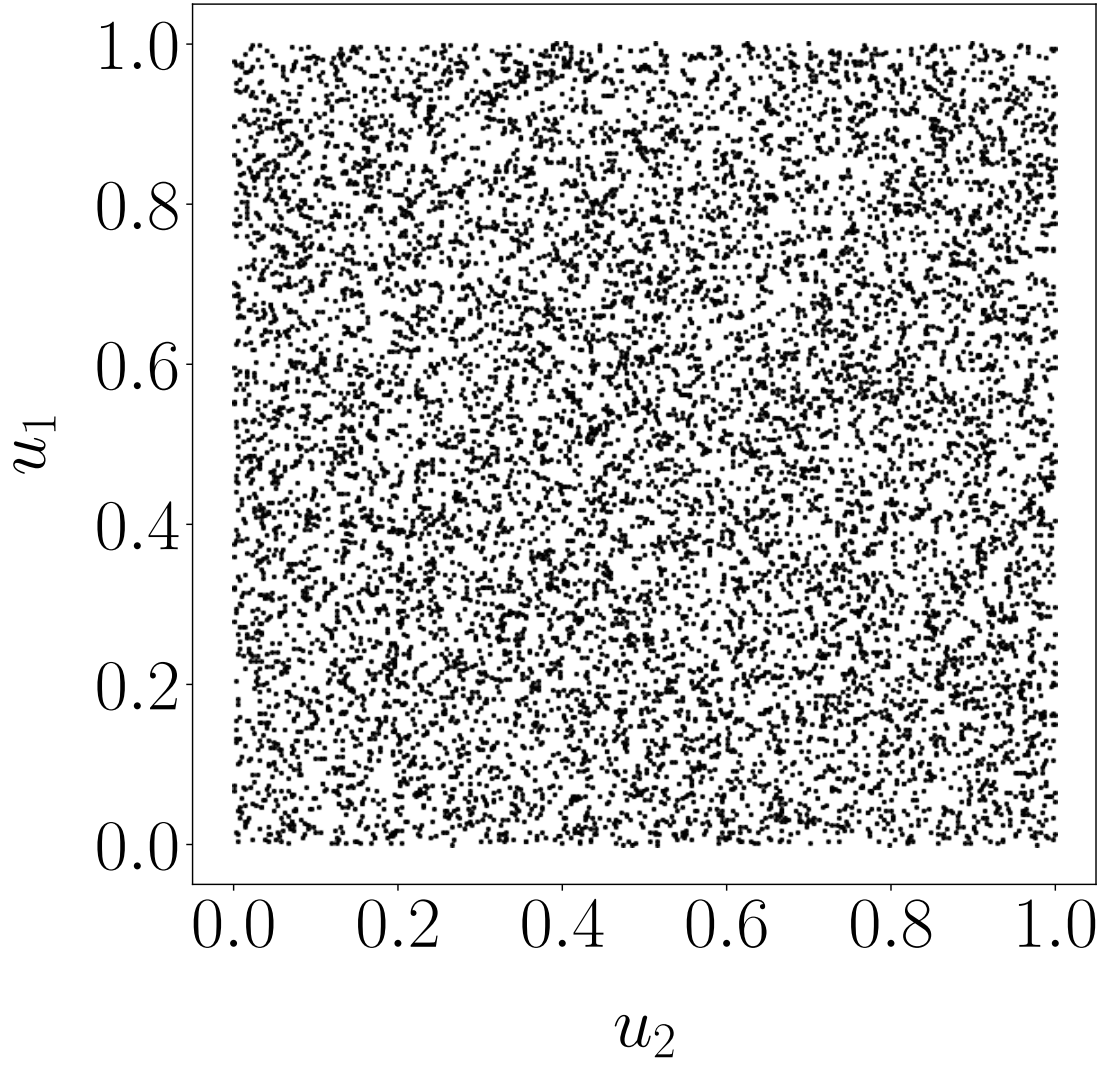
First idea

Obtain two uniform samples u_1 and u_2 and set:

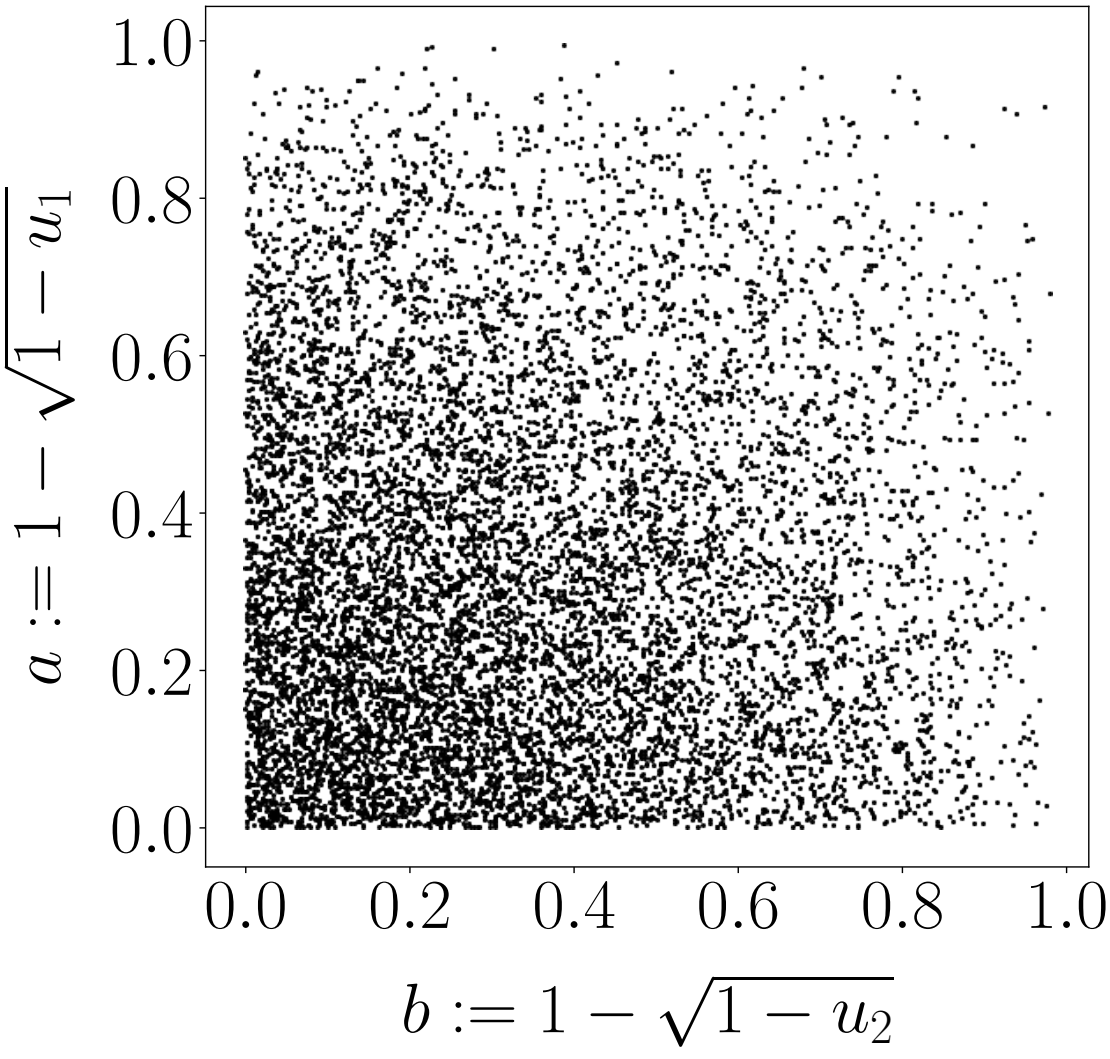
$$a := 1 - \sqrt{1 - u_1}$$

$$b := 1 - \sqrt{1 - u_2}$$

Uniform samples



First idea



What is the problem?

Marginal distributions are correct but joint distribution is wrong!

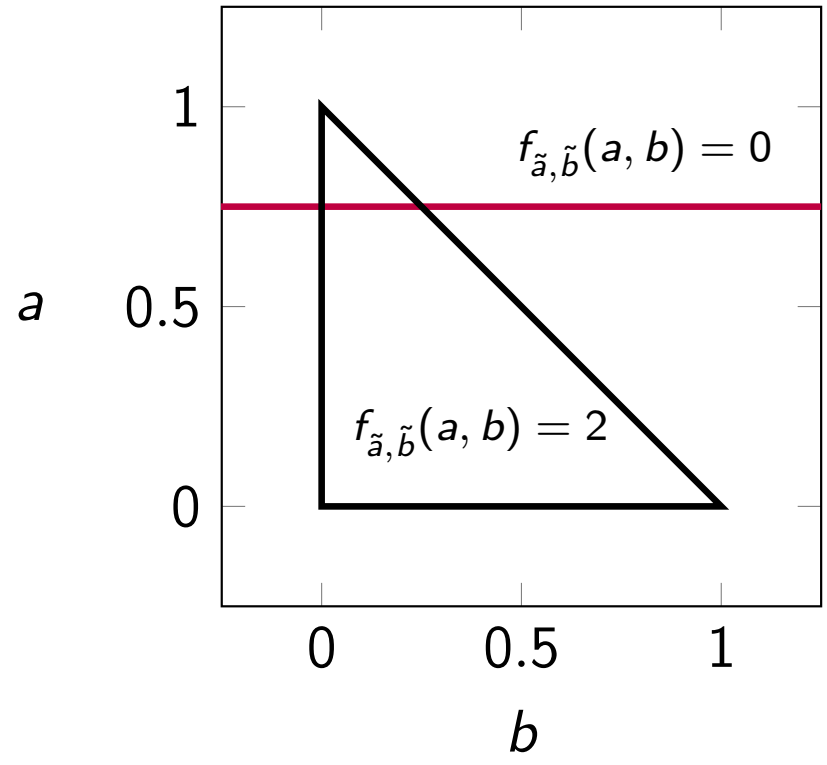
Samples of $\tilde{a}$ and $\tilde{b}$ are independent

We are sampling from the joint pdf

$$f_{\tilde{a}}(a) f_{\tilde{b}}(b) \quad \neq \quad f_{\tilde{a},\tilde{b}}(a,b) = f_{\tilde{a}}(a) f_{\tilde{b}\,|\,\tilde{a}}(b \,|\, a)$$

Solution: Sample from marginal and then from conditional distribution!

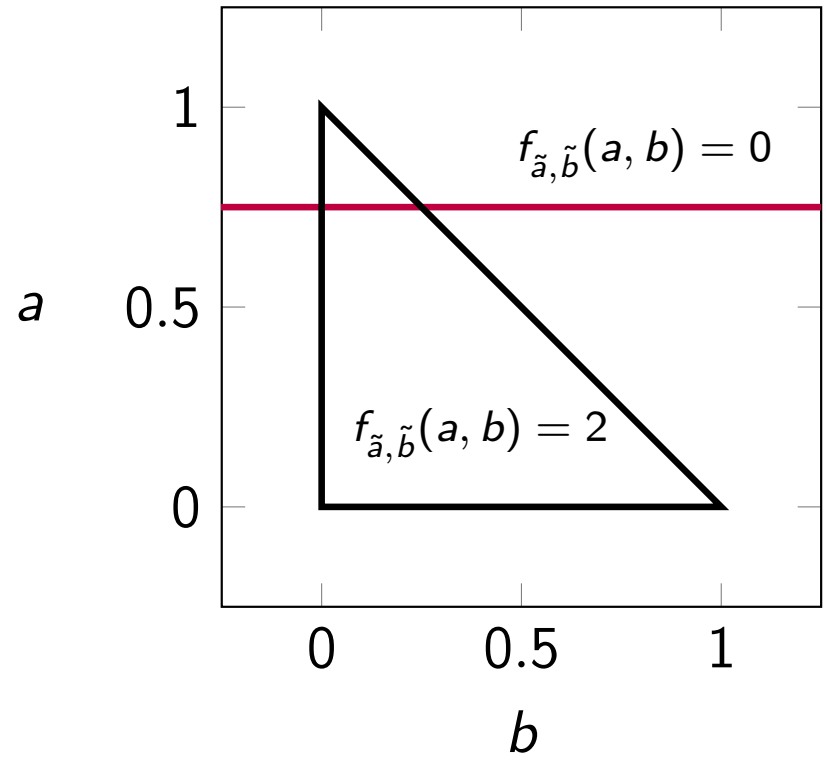
Conditional distribution of $\tilde{b}$ given $\tilde{a}$



$$f_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) = \frac{f_{\tilde{a},\tilde{b}}(a,b)}{f_{\tilde{a}}(a)} = \frac{1}{1-a} \quad b \in [0, 1-a]$$

$$F_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) = \int_0^b f_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a)\,\mathrm{d}b = \frac{b}{2(1-a)} \quad b \in [0, 1-a]$$

Conditional distribution of $\tilde{b}$ given $\tilde{a}$



$$F_{\tilde{b}\,|\,\tilde{a}}(b\,|\,a) = \frac{b}{1-a} \quad b \in [0, 1-a]$$

$$F^{-1}_{\tilde{b}\,|\,\tilde{a}}(u\,|\,a) = (1-a)u$$

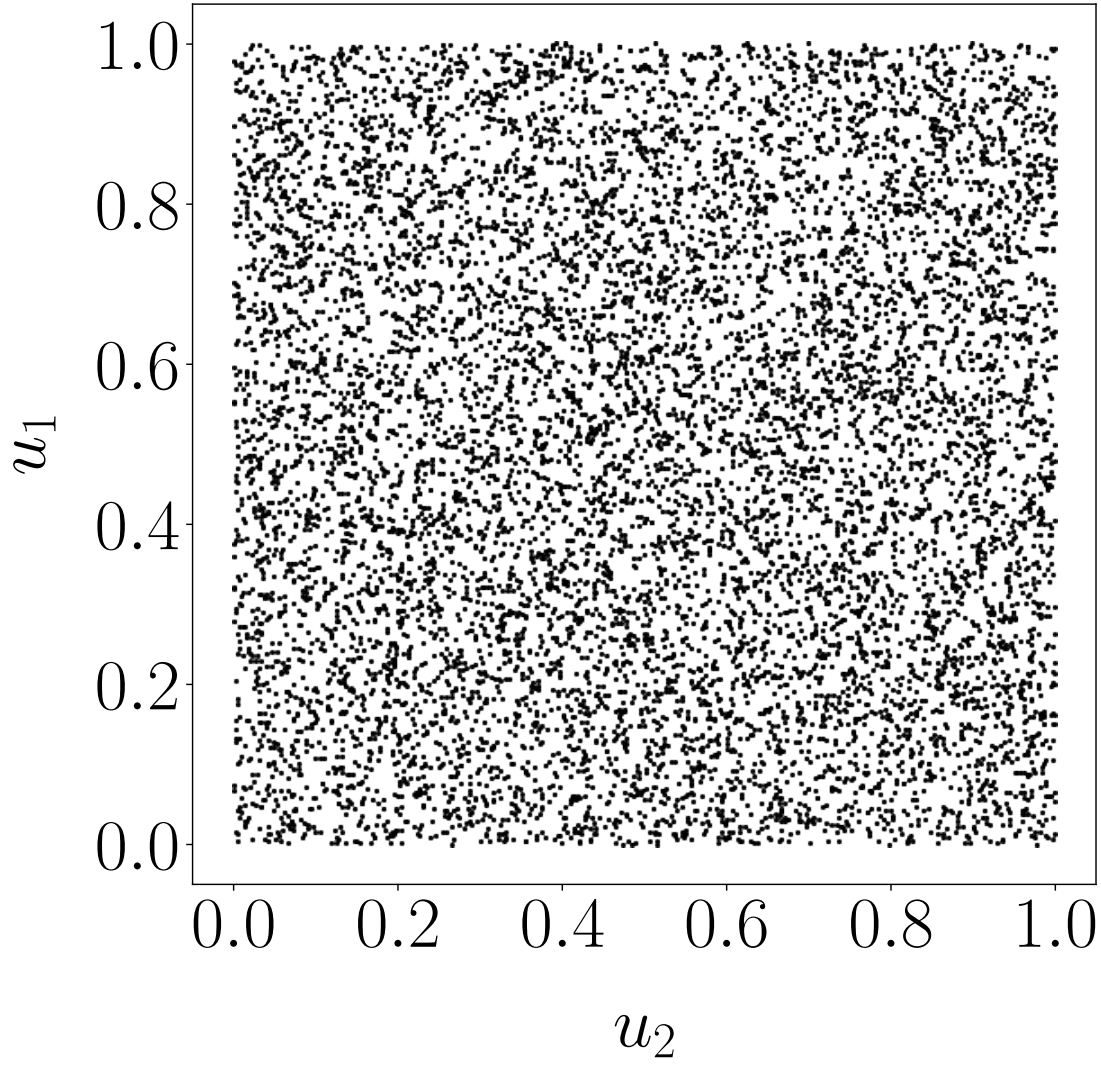
Second idea

Obtain two uniform samples u_1 and u_2 and set:

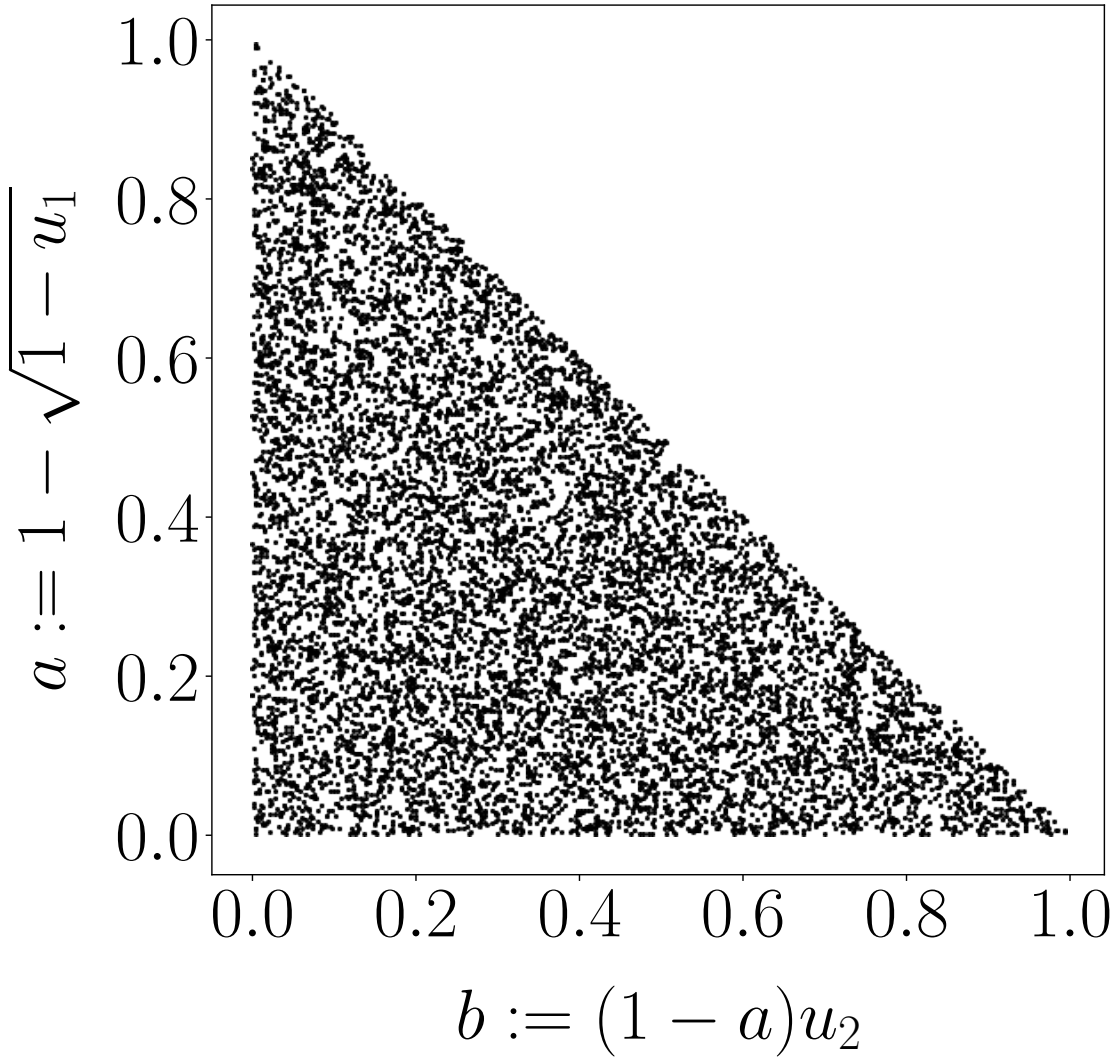
$$a := 1 - \sqrt{1 - u_1}$$

$$b := (1 - a)u_2$$

Uniform samples



Second idea



What have we learned?

How to simulate the joint distribution of multiple continuous random variables