Inverse Transform Sampling

Probability and Statistics for Data Science

Carlos Fernandez-Granda

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

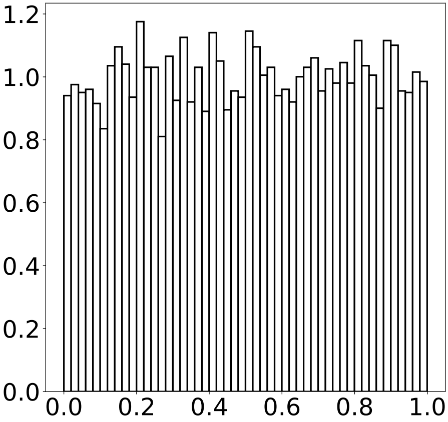
Motivation

Simulation is crucial for probabilistic modeling

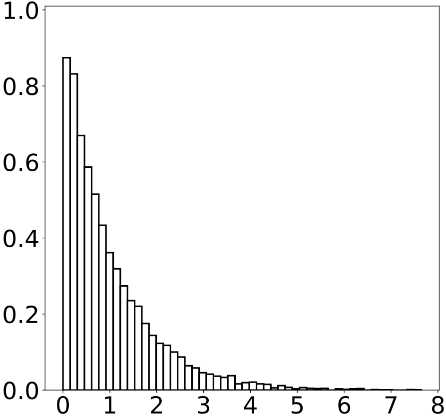
Strategy

1. Generate uniform samples in $[0, 1]$
2. Transform them so that they have the desired distribution

Histogram of uniform samples



Goal



Idea

Goal: Simulate random variable $\tilde{a}$

Input: Uniform random variable $\tilde{u}$

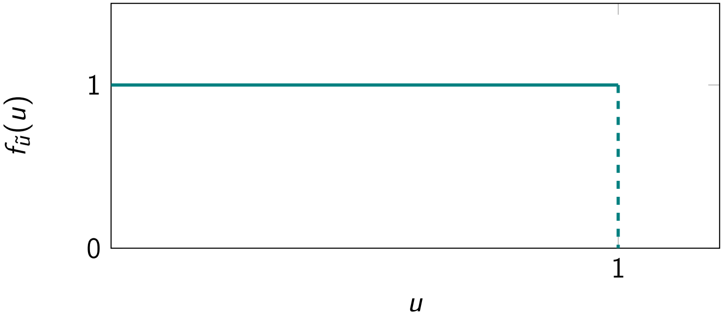
Strategy: Design g so that $g(\tilde{u})$ has the same distribution as $\tilde{a}$

For any interval $(x, y]$, we want

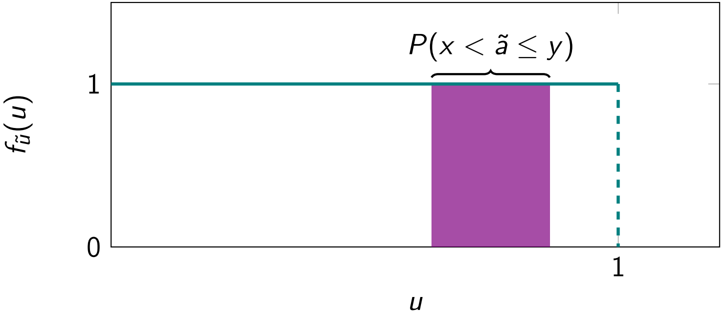
$$\mathrm{P}(x < g(\tilde{u}) \leq y) = \mathrm{P}(x < \tilde{a} \leq y)$$

Idea

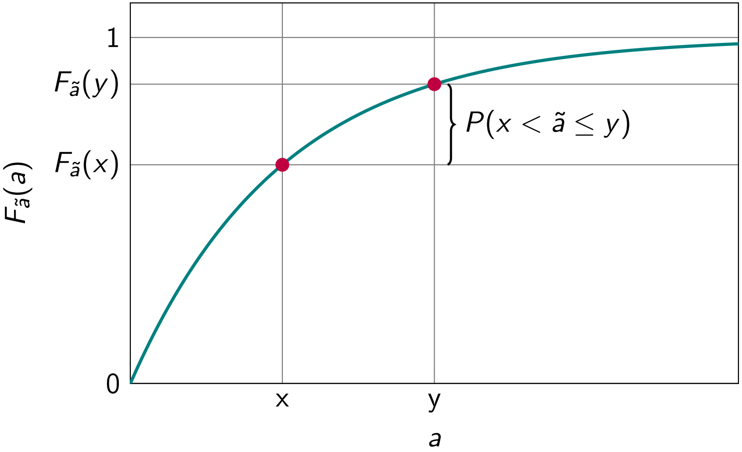
Length of set mapped to $(x, y]$?



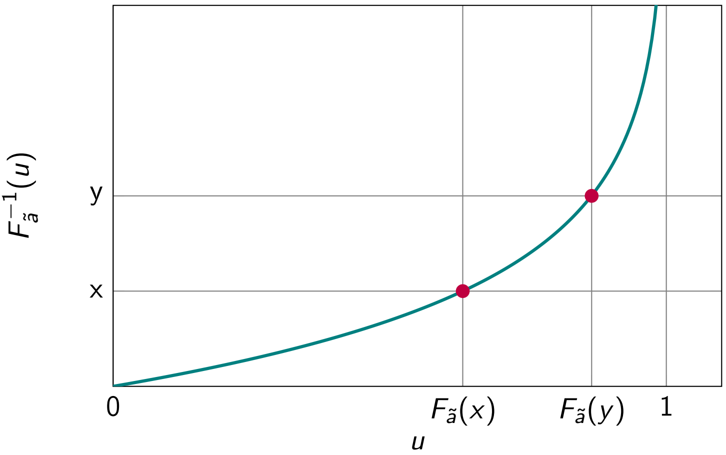
$\mathrm{P}(x < \tilde{a} \leq y)$



Remember the cdf?



Use the inverse cdf!



Inverse-transform sampling

Aim: Simulate sample from random variable $\tilde{a}$ with cdf $F_{\tilde{a}}$ using sample from uniform random variable $\tilde{u}$

Algorithm:

1. Obtain a sample u of $\tilde{u}$
2. Set $a := F_{\tilde{a}}^{-1}(u)$

Does this work?

$$
\begin{aligned}
F_{\tilde{b}}(y) &= \mathrm{P}\left(\tilde{b} \leq y\right) \\
&= \mathrm{P}\left(F_{\tilde{a}}^{-1}(\tilde{u}) \leq y\right) \\
&= \mathrm{P}\left(\tilde{u} \leq F_{\tilde{a}}(y)\right) \\
&= \int_{u=0}^{F_{\tilde{a}}(y)} \mathrm{d}u \\
&= F_{\tilde{a}}(y) \qquad \textit{Yes}!
\end{aligned}
$$

Exponential distribution

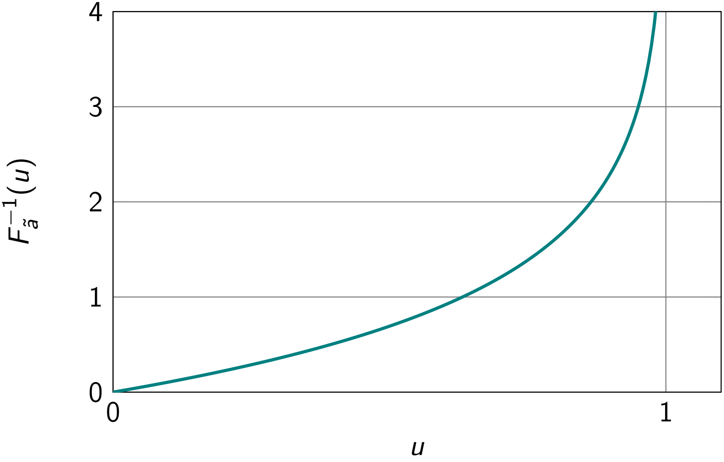
Aim: Sample from exponential random variable $\tilde{a}$ with parameter λ

$$F_{\tilde{a}}(a) := 1 - e^{-\lambda a}$$

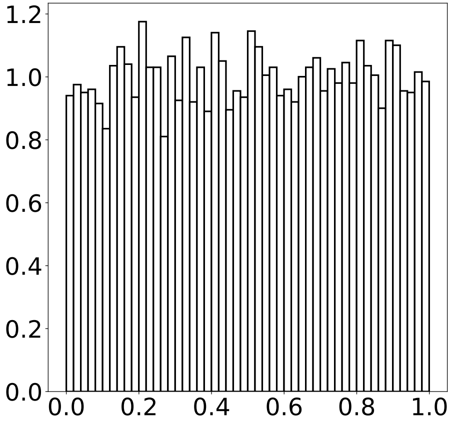
$$F_{\tilde{a}}^{-1}(u) = \frac{1}{\lambda} \log\left(\frac{1}{1-u}\right)$$

$F_{\tilde{a}}^{-1}(\tilde{u})$ is an exponential random variable with parameter λ

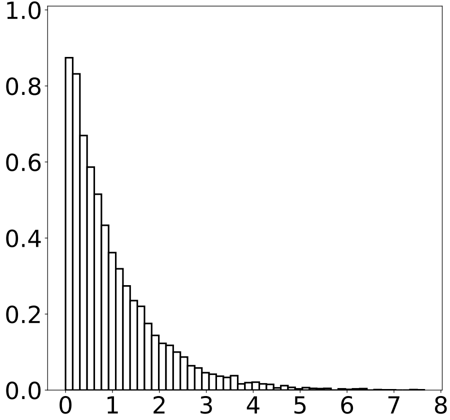
Exponential distribution



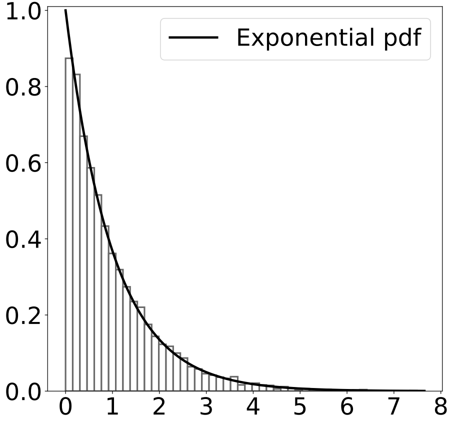
Histogram of uniform samples



Histogram of transformed samples



$\lambda_{\mathrm{ML}} = 0.9986$



What have we learned?

How to generate arbitrary continuous distributions from uniform samples