450D Political Methodology IV
TA Section 6

Vincent Bauer

November 3, 2017

Goal Today

- Transformation of Variables
- Hierarchical models

Road map to those goals:

1. Transformation of Variables (v2)
   - Motivation
   - Process
   - Derivation
   - Application: Inverse transform sampling
2. Multilevel models
   - Regression model examples in Stan
   - Common error messages
   - REML and `lmer`

Transformation of Variables

Motivation

Imagine we have some values X distributed according to a known distribution,

Motivation

Imagine we have some values X distributed according to a known distribution, such as Uniform[0,1].

Motivation

Imagine we have some values X distributed according to a known distribution, such as Uniform[0,1]. We write the probability density function as $f_x(x)$.

Motivation

Imagine we have some values X distributed according to a known distribution, such as Uniform[0,1]. We write the probability density function as $f_x(x)$.

$$f_x(x) = \begin{cases} 1, & \text{for } \ 0 < x < 1 \\ 0, & \text{for } \ x < 0, x > 1 \end{cases}$$

Suppose we then want to apply a transformation $g(x)$ to those values, defined as:

$$y = g(x) = \frac{1}{\lambda} - \log(x)$$

Motivation

Imagine we have some values X distributed according to a known distribution, such as Uniform[0,1]. We write the probability density function as $f_x(x)$.

$$f_x(x) = \begin{cases} 1, & \text{for } \; 0 < x < 1 \\ 0, & \text{for } \; x < 0, x > 1 \end{cases}$$

Suppose we then want to apply a transformation $g(x)$ to those values, defined as:

$$y = g(x) = \frac{1}{\lambda} - \log(x)$$

Q: What is the probability density function of the transformed values,

Motivation

Imagine we have some values X distributed according to a known distribution, such as Uniform[0,1]. We write the probability density function as $f_x(x)$.

$$f_x(x) = \begin{cases} 1, & \text{for} \ \ 0 < x < 1 \\ 0, & \text{for} \ \ x < 0, x > 1 \end{cases}$$

Suppose we then want to apply a transformation $g(x)$ to those values, defined as:

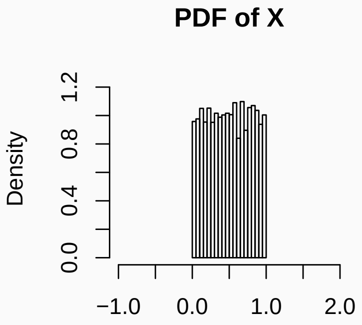
$$y = g(x) = \frac{1}{\lambda} - \log(x)$$

Q: What is the probability density function of the transformed values, that is, what is $f_y(y)$?

Simulation

```
m <- 10000
x <- runif(m, 0, 1)
f.y <- function(x, theta) -log(x)/theta

par(mfrow=c(1,2))
hist(x, freq=FALSE, main="PDF of X", ylim=c(0,1.25), xlim=c(-1,2))
f.y(x, theta=1) %>% hist(freq=FALSE, main="PDF of Y", xlab="y", ylim=c(0,1))
```



Process

We can answer this question using the tranformation of variables equation.

Process

We can answer this question using the tranformation of variables equation.

$$f_y(y) = f_x\{g^{-1}(y)\}\left|\frac{d}{dy}g^{-1}(y)\right|$$

Process

We can answer this question using the tranformation of variables equation.

$$f_y(y) = f_x\{g^{-1}(y)\}\left|\frac{d}{dy}g^{-1}(y)\right|$$

1. Solve for $g^{-1}(y)$, which is the inverse of $g(x)$

Process

We can answer this question using the tranformation of variables equation.

$$f_y(y) = f_x\{g^{-1}(y)\}\left|\frac{d}{dy}g^{-1}(y)\right|$$

1. Solve for $g^{-1}(y)$, which is the inverse of $g(x)$

$$\begin{aligned} y = g(x) &= \frac{1}{\lambda} - \log(x) \\ \lambda \cdot y &= -\log(x) \\ \exp(-\lambda \cdot y) &= x \\ &= g^{-1}(y) \end{aligned}$$

Process

2. Take the derivative of $g^{-1}(y)$ using the chain rule

Process

4. Take the derivative of $g^{-1}(y)$ using the chain rule

$$\begin{aligned}\frac{d}{dy}g^{-1}(y) &= \frac{d}{dy}\exp(-\lambda \cdot y) \\ &= \exp(-\lambda \cdot y) \cdot -\lambda\end{aligned}$$

Process

6. Take the derivative of $g^{-1}(y)$ using the chain rule

$$\begin{aligned}\frac{d}{dy}g^{-1}(y) &= \frac{d}{dy}\exp(-\lambda \cdot y)\\ &= \exp(-\lambda \cdot y) \cdot -\lambda\end{aligned}$$

7. Solve for $f_x(g^{-1}(x))$

$$\begin{aligned}f_x(x) &= 1 \quad \text{for } 0 < x < 1\\ f_x(g^{-1}(x)) &= 1 \quad \text{for } 0 < g^{-1}(x) < 1\\ &\qquad\quad 0 < g^{-1}(x) < 1\\ &\qquad\quad 0 < \exp(-\lambda \cdot y) < 1\end{aligned}$$

Process

8. Combine these pieces of information

$$f_y(y) = f_x\{g^{-1}(y)\}\left|\frac{d}{dy}g^{-1}(y)\right|$$

so

$$f_y(y) = \begin{cases} 1 \times |\exp(-\lambda \cdot y) \cdot -\lambda|, & \text{for } \; 0 < y < \infty \\ 0 \times |\exp(-\lambda \cdot y) \cdot -\lambda|, & \text{for } \; y < 0 \end{cases}$$

Process

8. Combine these pieces of information

$$f_y(y) = f_x\{g^{-1}(y)\}\left|\frac{d}{dy}g^{-1}(y)\right|$$

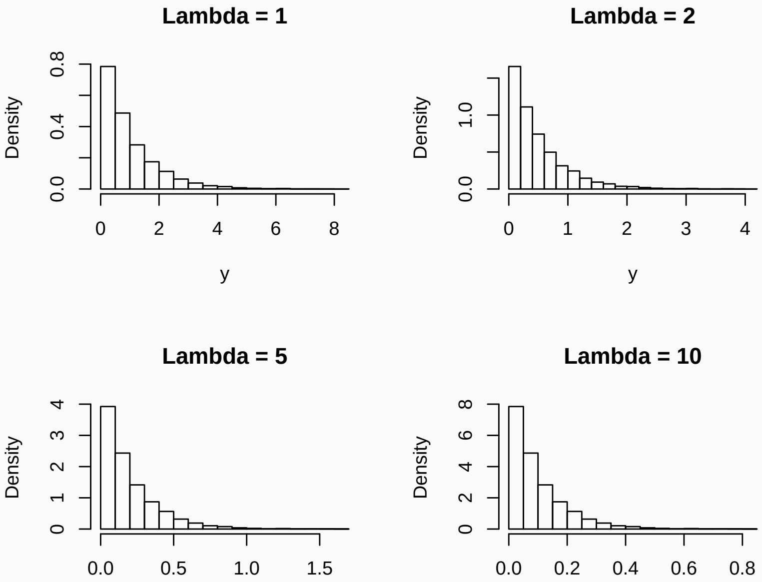
so

$$f_y(y) = \begin{cases} 1 \times |\exp(-\lambda \cdot y) \cdot -\lambda|, & \text{for } \ 0 < y < \infty \\ 0 \times |\exp(-\lambda \cdot y) \cdot -\lambda|, & \text{for } \ y < 0 \end{cases}$$

It turns out that this is the PDF of the exponential distribution.

$$Exponential(\theta) = \lambda \cdot \exp(-\lambda \cdot x)$$

Simulation



Lambda = 1
Density
0.0
0.4
0.8
0
2
4
6
8
y
Lambda = 2
Density
0.0
1.0
0
1
2
3
4
y
Lambda = 5
Density
0
1
2
3
4
0.0
0.5
1.0
1.5
Lambda = 10
Density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8

Derivation

Where does this weird formula come from?

Derivation

Where does this weird formula come from? We can derive this result if we approach the problem from the CDFs of the variables.

Derivation

Where does this weird formula come from? We can derive this result if we approach the problem from the CDFs of the variables.

Start with

$$y = g(x)$$

Derivation

Where does this weird formula come from? We can derive this result if we approach the problem from the CDFs of the variables.

Start with

$$y = g(x)$$

Then write:

Derivation

Where does this weird formula come from? We can derive this result if we approach the problem from the CDFs of the variables.

Start with

$$y = g(x)$$

Then write:

$$\begin{aligned} CDF_y(y) &= F_y(y) \\ &= P(Y \leq y) \\ &= P\left(g(x) \leq y\right) \\ &= P(x \leq g^{-1}(y)) \\ &= F_x(g^{-1}(y)) \\ &= CDF_x(g^{-1}(y)) \end{aligned}$$

Derivation

We found that

$$CDF_y(y) = CDF_x(g^{-1}(y))$$

Derivation

We found that

$$CDF_y(y) = CDF_x(g^{-1}(y))$$

But we want the PDF not the CDF, so take the derivative

Derivation

We found that

$$CDF_y(y) = CDF_x(g^{-1}(y))$$

But we want the PDF not the CDF, so take the derivative

$$\begin{aligned} PDF_y(y) &= \frac{d}{dy} CDF_y(y) \\ &= \frac{d}{dy} CDF_x(g^{-1}(y)) \end{aligned}$$

Using the chain-rule

$$= f_x(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

Derivation

We found that

$$CDF_y(y) = CDF_x(g^{-1}(y))$$

But we want the PDF not the CDF, so take the derivative

$$\begin{aligned} PDF_y(y) &= \frac{d}{dy} CDF_y(y) \\ &= \frac{d}{dy} CDF_x(g^{-1}(y)) \end{aligned}$$

Using the chain-rule

$$= f_x(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

And then you need the absolute value for some reason.

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

If $U \sim Uniform[0, 1]$, then $X = F^{-1}(U)$ has the distribution F, where F is the CDF of a distribution and F^{-1} is its quantile function (inverse-CDF).

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

If $U \sim Uniform[0,1]$, then $X = F^{-1}(U)$ has the distribution F, where F is the CDF of a distribution and F^{-1} is its quantile function (inverse-CDF).

In other words, we can throw a random uniform [0,1] sample into the inverse CDF and the results will be distributed according to the CDF.

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

If $U \sim Uniform[0, 1]$, then $X = F^{-1}(U)$ has the distribution F, where F is the CDF of a distribution and F^{-1} is its quantile function (inverse-CDF).

In other words, we can throw a random uniform [0,1] sample into the inverse CDF and the results will be distributed according to the CDF.

This method is very fast if we have the quantile functions,

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

If $U \sim \mathit{Uniform}[0,1]$, then $X = F^{-1}(U)$ has the distribution F, where F is the CDF of a distribution and F^{-1} is its quantile function (inverse-CDF).

In other words, we can throw a random uniform [0,1] sample into the inverse CDF and the results will be distributed according to the CDF.

This method is very fast if we have the quantile functions, but there are some distributions that we can't analytically solve the quantile function.

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

If $U \sim \mathit{Uniform}[0,1]$, then $X = F^{-1}(U)$ has the distribution F, where F is the CDF of a distribution and F^{-1} is its quantile function (inverse-CDF).

In other words, we can throw a random uniform [0,1] sample into the inverse CDF and the results will be distributed according to the CDF.

This method is very fast if we have the quantile functions, but there are some distributions that we can't analytically solve the quantile function. The normal distribution is one of these.

Application: Inverse Transformation Sampling

We can use this general idea to sample points from any distribution for which we know the inverse-CDF (quantile function).

If $U \sim Uniform[0,1]$, then $X = F^{-1}(U)$ has the distribution F, where F is the CDF of a distribution and F^{-1} is its quantile function (inverse-CDF).

In other words, we can throw a random uniform [0,1] sample into the inverse CDF and the results will be distributed according to the CDF.

This method is very fast if we have the quantile functions, but there are some distributions that we can't analytically solve the quantile function. The normal distribution is one of these. But we can approximate the quantile function with a polynomial, which is

Intuition

From wikipedia:

We are randomly choosing a proportion of the area under the curve and returning the number in the domain such that exactly this proportion of the area occurs to the left of that number.

Intuition

From wikipedia:

We are randomly choosing a proportion of the area under the curve and returning the number in the domain such that exactly this proportion of the area occurs to the left of that number.

Intuitively, we are unlikely to choose a number in the far end of tails because there is very little area in them which would require choosing a number very close to zero or one.

Process

Start with $U \sim Unif(0,1)$ generated by a RNG, and the CDF of the exponential distribution, which is $F(x) = 1 - \exp(-\lambda x)$.

Process

Start with $U \sim Unif(0, 1)$ generated by a RNG, and the CDF of the exponential distribution, which is $F(x) = 1 - \exp(-\lambda x)$.

Then, obtain the inversion function by solving $y = F(x)$:

Process

Start with $U \sim Unif(0, 1)$ generated by a RNG, and the CDF of the exponential distribution, which is $F(x) = 1 - \exp(-\lambda x)$.

Then, obtain the inversion function by solving $y = F(x)$:

$$\begin{aligned} y &= 1 - \exp(-\lambda x) \\ y - 1 &= -\exp(-\lambda x) \\ \log(y - 1) &= -\lambda x \\ \frac{1}{-\lambda} \log(y - 1) &= x \end{aligned}$$

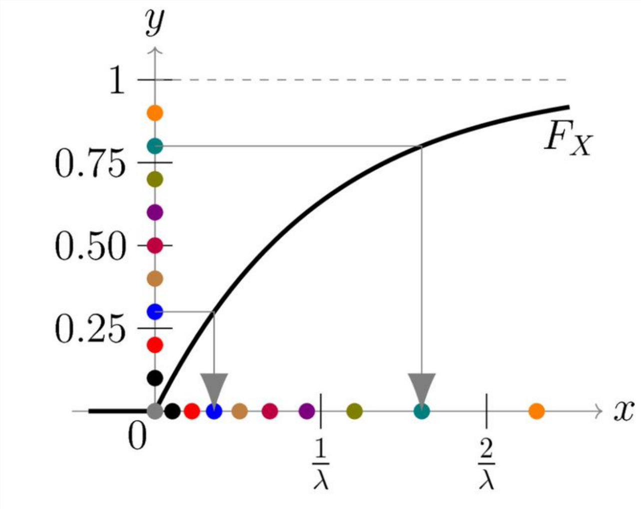
Process

Start with $U \sim \mathit{Unif}(0,1)$ generated by a RNG, and the CDF of the exponential distribution, which is $F(x) = 1 - \exp(-\lambda x)$.

Then, obtain the inversion function by solving $y = F(x)$:

$$
\begin{aligned}
y &= 1 - \exp(-\lambda x) \\
y - 1 &= -\exp(-\lambda x) \\
\log(y - 1) &= -\lambda x \\
\frac{1}{-\lambda}\log(y - 1) &= x
\end{aligned}
$$

Plug our values of U into $F^{-1}(y)$, and get output $\sim F(y)$.



Notice that even though our y axis is uniformly distributed, our x axis has many small values and a few large values, the exponential distribution.

Derivation

Claims: if U is a uniform random variable on (0,1), then $F^{-1}(U) \sim F$.

$$\begin{aligned} Pr(F^{-1}(U) \leq x) &= Pr(F(F^{-1}(U)) \leq F(x)) \\ &= Pr(U \leq F(x)) \end{aligned}$$

By properties of the uniform dist on (0,1), $Pr(U \leq x) = x$

$$= F(x)$$

Multilevel models

Multilevel models

*on the whiteboard