Functions of a Random Variable

Most often in inferential statistics, we are not interested in a single variable.

We are usually interested in multiple variables, and more importantly,
how one variable is functionally related to other variables

Functions of a Random Variable

Most often in inferential statistics, we are not interested in a single variable.

We are usually interested in multiple variables, and more importantly,
how one variable is functionally related to other variables

So pay attention here!

Linear Functions

$Y = aX + b$

We can think of b as a location parameter

We can thing of a as a scale parameter

Linear Functions

$Y = aX + b$
We can think of b as a location parameter
We can thing of a as a scale parameter

$$
\begin{aligned}
F_Y(y) = &\quad P(Y \leq y) && \text{By definition of F} \\
= &\quad P(aX + b < y) && \text{By definition of Y} \\
= &\quad P(X \leq \tfrac{y-b}{a}) && \text{a little algebra} \\
= &\quad F_X(\tfrac{y-b}{a}) && \text{By definition of F}
\end{aligned}
$$

Linear Functions

$Y = aX + b$
We can think of b as a location parameter
We can thing of a as a scale parameter

$$\begin{aligned} F_Y(y) = & \quad P(Y \leq y) && \text{By definition of F} \\ = & \quad P(aX + b < y) && \text{By definition of Y} \\ = & \quad P(X \leq \tfrac{y-b}{a}) && \text{a little algebra} \\ = & \quad F_X(\tfrac{y-b}{a}) && \text{By definition of F} \end{aligned}$$

The CDF of Y and X are related.

Notice the form: 1)Subtract the location 2) Divide by scale

Gaussian Application

I have a Gaussian variable X with mean (location) μ and standard error (scale) σ.
What is P(a<X<b)?
Well, we already known that $P(a < X < b) = F_X(b) - F_X(a)$
Consider the linear functions

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$
$$X = \sigma Z + \mu$$

Now Z is standard normal, i.e. N(0,1)

App. Continued

$$
\begin{aligned}
F_X(a) &= P(X < a) && \text{Definition of } F_X \\
&= P(\sigma Z + \mu < a) && \text{Definition of Z} \\
&= P(Z < \tfrac{a-\mu}{\sigma}) && \text{a little algebra} \\
&= \boxed{\Phi\left(\tfrac{a-\mu}{\sigma}\right)}
\end{aligned}
$$

Significance: X is normal but not standard normal.
We can evaluate $F_X(a)$ if we know $\Phi\left(\frac{a-\mu}{\sigma}\right)$
Return: What is $P(a < X < b)$?

$$\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Example

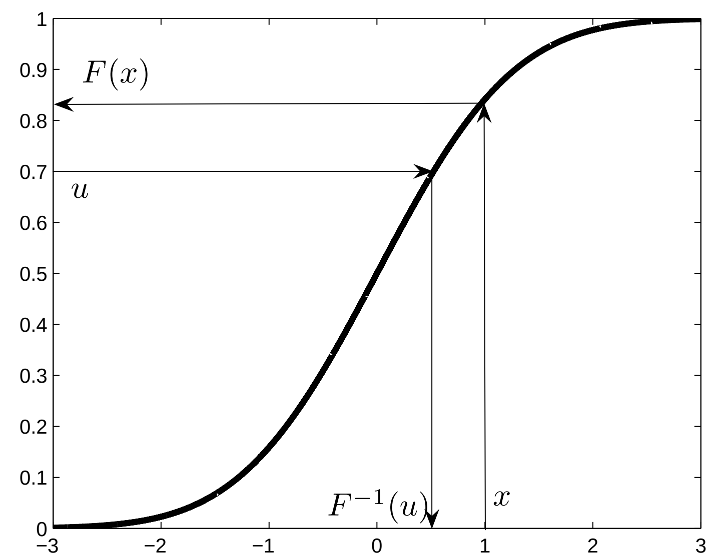
Suppose X is distributed N(4,25) (N(μ, σ^2)).
What is $P(X \leq 5)$?

$$
\begin{aligned}
P(X \leq 5) &= \Phi\left(\frac{5-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{5-4}{5}\right) \\
&= \Phi(.2) = .5793
\end{aligned}
$$

What is $P(X > 5)$?

$$
\begin{aligned}
P(X > 5) &= 1 - P(X \leq 5) \\
&= 1 - .5793 = .4207
\end{aligned}
$$

Simulation Example



If $u \leq F(x)$ then $F^{-1}(u) \leq x$
Make u a uniform random number U
If $U \leq F(x)$ then $F^{-1}(U) \leq x$

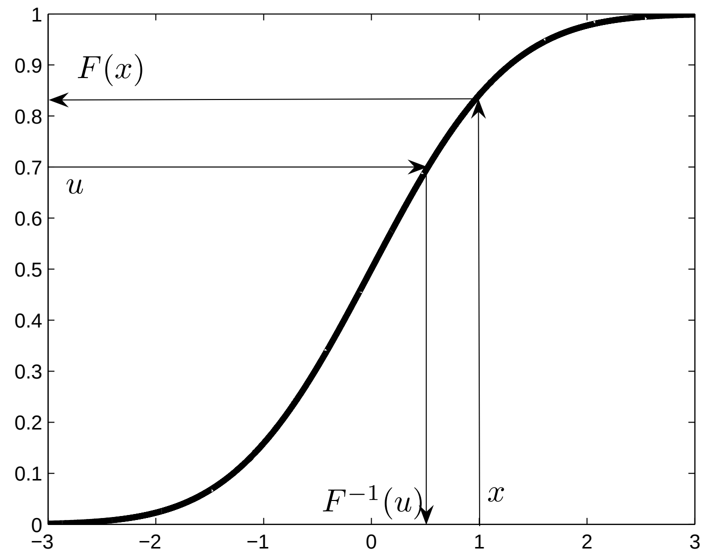
Example cont.

If $U \leq F(x)$ then $F^{-1}(U) \leq x$
$F(x)$ is a number... call it b

$$
\begin{aligned}
P(U \leq b) &= b && \text{by Def. of Uniform RV} \\
P(U \leq F(x)) &= F(x) && \text{Substitute} F(x) = b \\
P(F^{-1}(U) \leq x) &= F(x) && \text{From first line on page}
\end{aligned}
$$

In words:
The random variable $F^{-1}(U)$ has the distribution $F(x)$

Simulating RVs

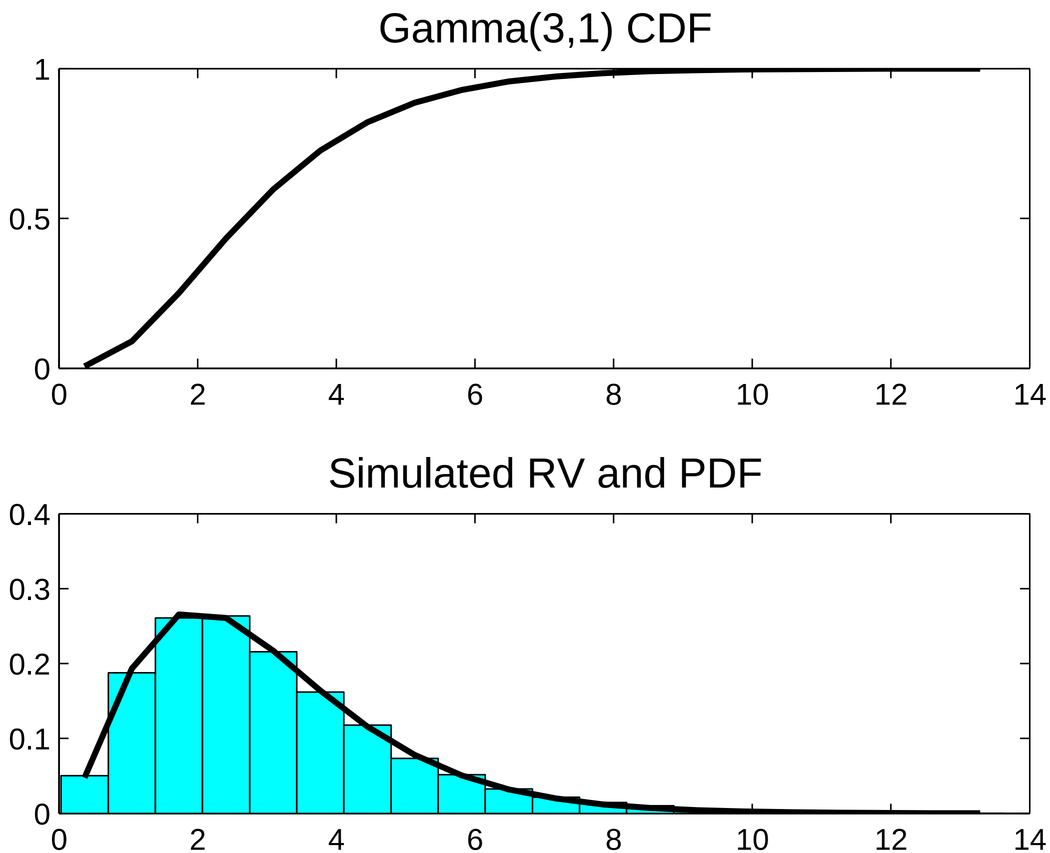


To simulate 100 RVs from $F(x)$:

1. Simulate 100 uniform RVs u
2. $X = F^{-1}(u)$

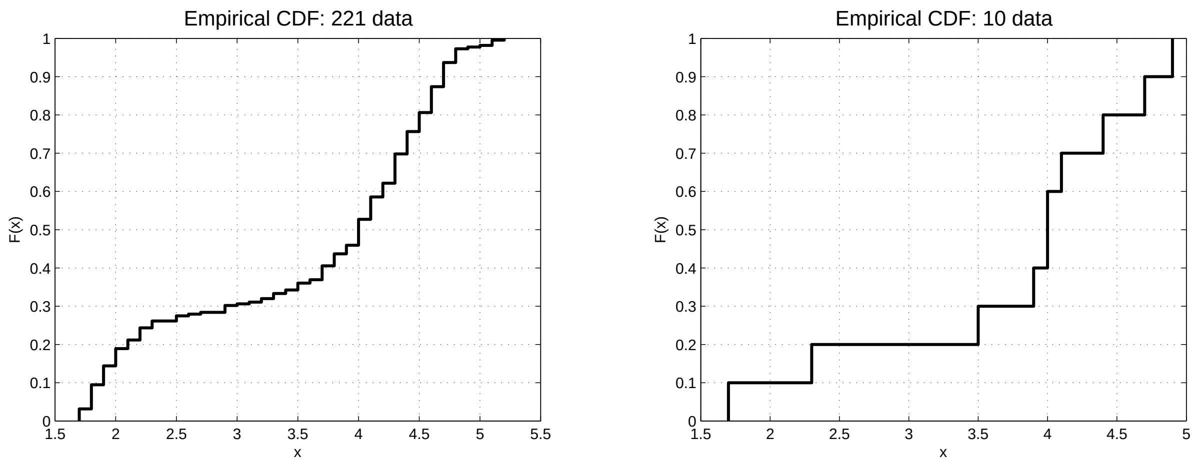
Simulation Example

Simulation from Gamma(3,1)



Bootstrap Example

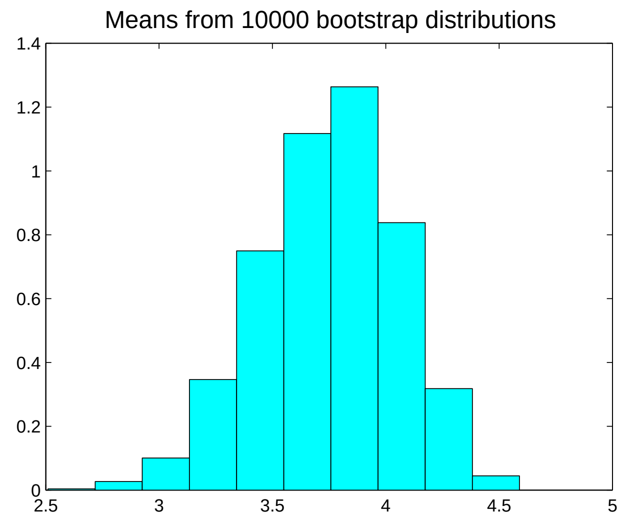
Durations of Old Faithful eruptions



What is the mean?
Mean of 221 data is 3.58
Mean of 10 data is 3.75
How can we make an interval?
What distribution does it come from?
Simulate data sets from empirical CDF.

Bootstrap Distribution

Generate many ”duration” datasets from ECDF
For each dataset, find the mean



2.5% is below 3.18   97.5% is below 4.16
95% is between 3.18 and 4.16
We are 95% confident the mean is between 3.18 and 4.16... all this from just 10 of the 221 data.
Mean of 221 data was 3.58