Statistical Inference Overview

Jeffrey Leek
Johns Hopkins Bloomberg School of Public Health

Statistical Inference Content

- Basic probability
- Likelihood
- Common distributions
- Asymptotics
- Confidence intervals
- Hypothesis tests
- Power
- Bootstrapping
- Non-parametric tests
- Basic bayesian statistics

Example

Suppose that the proportion of help calls that get addressed in a random day by a help line is given by

$$f(x) = \begin{cases} 2x & \text{for } 1 > x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Is this a mathematically valid density?

The normal distribution

- A random variable is said to follow a **normal** or **Gaussian** distribution with mean μ and variance σ^2 if the associated density is

$$(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/2\sigma^2}$$

  If X a RV with this density then $E[X] = \mu$ and $Var(X) = \sigma^2$

- We write $X \sim N(\mu, \sigma^2)$
- When $\mu = 0$ and $\sigma = 1$ the resulting distribution is called **the standard normal distribution**
- The standard normal density function is labeled ϕ
- Standard normal RVs are often labeled Z

Example bootstrap code

```
B <- 1000
n <- length(gmVol)
resamples <- matrix(sample(gmVol,
                           n * B,
                           replace = TRUE),
                    B, n)
medians <- apply(resamples, 1, median)
sd(medians)
[1] 3.148706
quantile(medians, c(.025, .975))
    2.5%    97.5%
582.6384 595.3553
```