Sampling Distributions

Stat 250

Click here for PDF version

Roadmap

Inferential statistics relies on understanding the distribution of the statistic of interest

1. Permutation tests: estimated the distribution of the test statistic by making H_0 true and then using permutation resampling (Chapter 3)
2. Estimation: estimate the distribution of a statistic without assuming a specific parameter value

- Appeal to probability (Chapter 4)
- Simulate the sampling process from the population (Chapter 4)
- Simulate the process using *bootstrap resampling* (Chapter 5)

Example

Let $T = \sum_{i=1}^{n} X_i$ where $X_i \overset{iid}{\sim} \text{Exp}(\lambda)$. Find the sampling distribution of T.

Recall:

- PDF of $X \sim \text{Exp}(\lambda)$: $f(x) = \lambda e^{-\lambda x}, x \geq 0$
- Moment generating function: $M_X(t) = \dfrac{\lambda}{\lambda - t}$

Example

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\mathrm{Exp}(\lambda)$ with PDF $f(x) = \lambda e^{-\lambda}, \lambda > 0, x > 0$.

Find the PDF of $X_{min} = \min(X_1, X_2, \ldots, X_n)$.

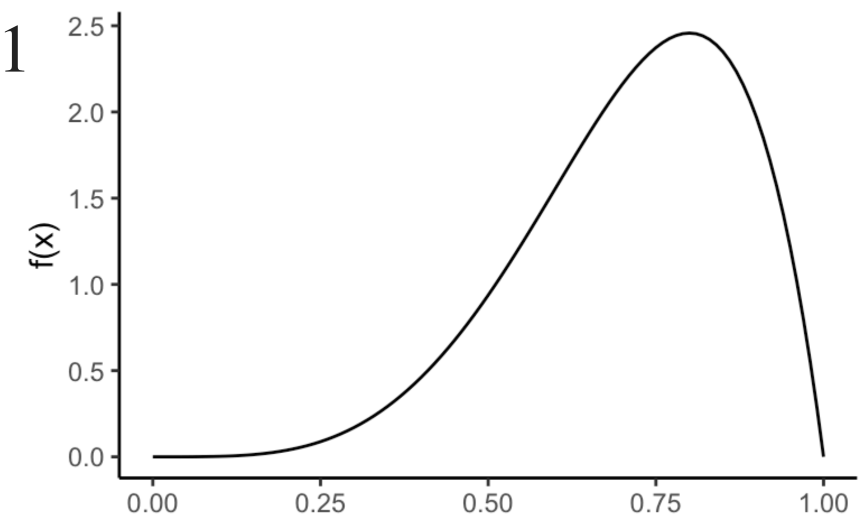
Simulation example

Let $T = \text{median}\{X_1, X_2, \ldots, X_{10}\}$ where $X_i \overset{iid}{\sim} \text{Beta}(5, 2)$.

Beta PDF

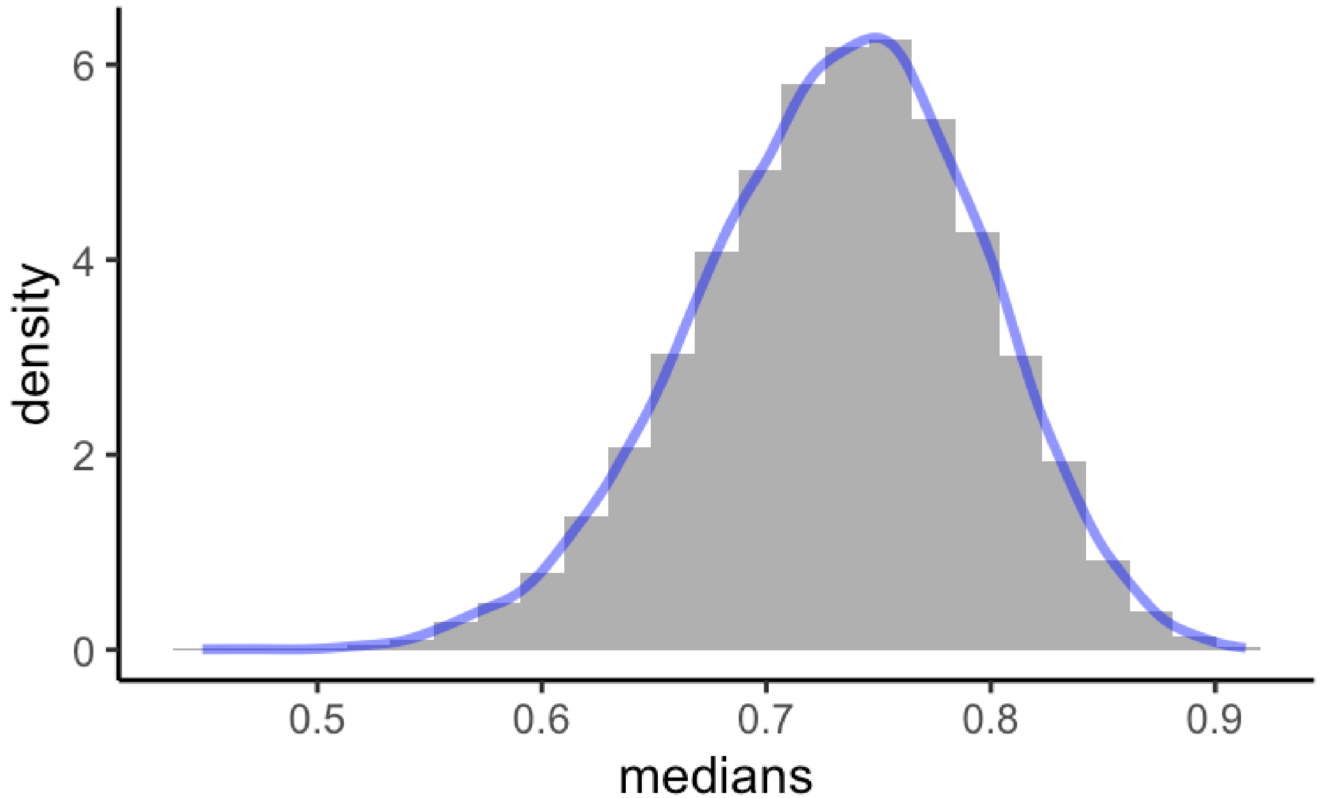
$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}$$

$$0 \leq x \leq 1; \alpha > 0, \beta > 0$$



Simulation example

- Draw 10,000 samples of size $n = 10$ from Beta(5, 2)
- Calculate the median for each simulated sample



Simulation example

```r
# Set seed for reproducibility
set.seed(492017)

N <- 10^4
medians <- numeric(N)

for(i in seq_len(N)) {
  # Draw sample of size 10 from Beta(5, 2)
  draws <- rbeta(n = 10, shape1 = 5, shape2 = 2)

  # Calculate + store the median
  medians[i] <- median(draws)
}
```

Simulation example

We can calculate the mean, standard error, and probabilities from simulations

```
# Mean of sampling distribution
mean(medians)
## [1] 0.7318108

# Standard error of sampling distribution
sd(medians)
## [1] 0.06300146

# Probability of an event from sampling distribution
mean(medians <= 0.6)
## [1] 0.0247
```

Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. samples from a χ^2_3 distribution.

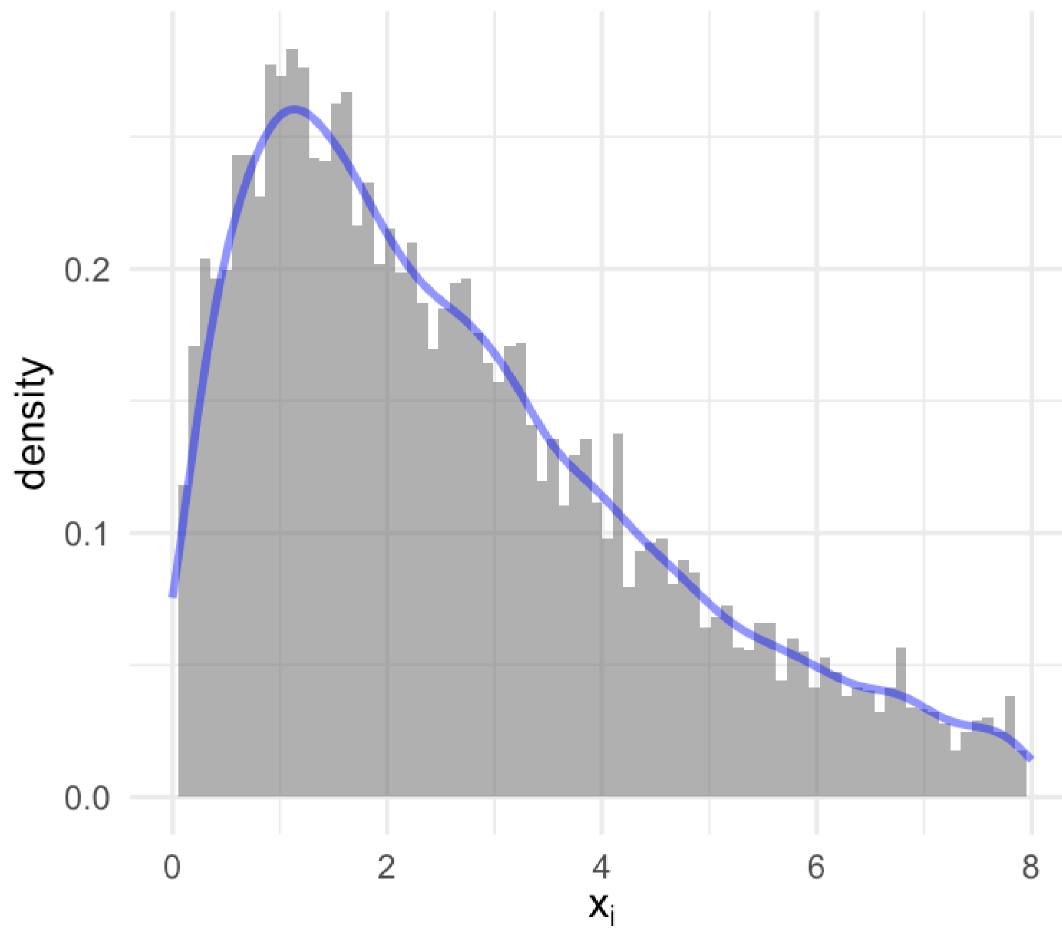
Population

n = 10

n = 20

n = 40

n = 80



Continuity correction

Let $X_1, \dots, X_n$ be i.i.d. samples from a $\text{Bin}(n = 10, p = 0.75)$ distribution.

Population

n = 10

n = 20

n = 40

n = 80