Lecture 4: Evaluating estimators

STA238: Probability, Statistics, and Data Analysis II

Fred Song

Week of July 15, 2024 Lec 4

Example: Estimating the number of German tanks

- The Allied forces during WWII wanted to estimate the total number of German tanks.
- They used the serial numbers of captured tanks to estimate the total number of tanks produced.

See Sections 20.1 and 1.5 from Dekking *et al.*

Example: Modelling the serial numbers

- Denote serial numbers of n captured German tanks (recoded as integer values) with x_1, x_2, ..., x_n.
- These serial numbers are from 1, 2, ..., θ, $\theta \geq n$.
- Assume the tanks are captured randomly; each tank has an have equal likelihood of being captured.
- Thus, we can model serial number i as a realization of X_i with

$$P(X_i = x) = \frac{1}{\theta}$$

for $x = 1, 2, \ldots, \theta$ and $i = 1, 2, \ldots, n$.

$X_1, X_2, \ldots, X_n$ are

____

but not _________________________.

Example: Modelling the serial numbers

- Denote serial numbers of n captured German tanks (recoded as integer values) with x_1, x_2, ..., x_n.
- These serial numbers are from 1, 2, ..., θ, $\theta \geq n$.
- Assume the tanks are captured randomly; each tank has an have equal likelihood of being captured.
- Thus, we can model serial number i as a realization of X_i with

$$P(X_i = x) = \frac{1}{\theta}$$

for $x = 1, 2, \ldots, \theta$ and $i = 1, 2, \ldots, n$.

$X_1, X_2, \ldots, X_n$ are

____

but not ________________________.

$$P(X_1 = x_1 | X_2 = x_2) \neq P(X_1 = x_1)$$

An unbiased estimator based on the sample mean, T_1

What is the expected value of $\bar{X}_n = \sum_{i=1}^{n} X_i / n$?

An unbiased estimator based on the sample mean, T_1

What is the expected value of $\bar{X}_n = \sum_{i=1}^{n} X_i \big/ n$?

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1) = \frac{1}{\theta}\sum_{i=1}^{\theta} i = \frac{\theta+1}{2}$$

$$\implies T_1 = 2\bar{X}_n - 1$$

is an unbiased estimator for θ.

Method of moments

The kth **moment** of a random variable X is defined by $\mu'_k = \mathbb{E}\left[X^k\right]$.

The corresponding kth **sample moment** is defined by $m_k = \sum_{i=1}^{n} X_i^k \big/ n$ for a sample X_1, X_2, ..., X_n that are identically distributed as X.

The **method of moments** constructs an estimator for a parameter $\theta = g\left(\mu'_1, \mu'_2, \ldots, \mu'_K\right)$ as a function of the sample moments $\hat{\theta} = g((m_1, m_2, \ldots, m_K)$.

- For example, the (sample) mean is the first (sample) moment.
- Sample moments are unbiased estimators of the population moments if they exist.
- Independent samples generally produces consistent method of moment estimators.

Exercises

Suppose $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} F$. For each of the following F, construct the method of moment estimator for θ.

1. $\text{Unif}(0, \theta)$
2. $\text{Geom}(p)$ with pmf $p(x) = (1-p)^x p$
3. $\mathcal{N}(0, \theta)$

Exercises

Suppose $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} F$. For each of the following F, construct the method of moment estimator for θ.

1. $\text{Unif}(0, \theta)$
2. $\text{Geom}(p)$ with pmf $p(x) = (1-p)^x p$
3. $\mathcal{N}(0, \theta)$

- Method of moments is a way to construct estimators based on the law of large numbers.
- The method is simple and intuitive.
- The method does not depend on the model distribution.
- Method of moments estimators don't always yield unbiased and sensible estimates.

An unbiased estimator based on the maximum, T_2

$$P(X_{(n)} = x) = \binom{x-1}{n-1} \Big/ \binom{\theta}{n}$$

$$= \underline{\hspace{3cm}}$$

What is the expected value of $X_{(n)}$?

An unbiased estimator based on the maximum, T_2

What is the expected value of $X_{(n)}$?

$$P(X_{(n)} = x) = \binom{x-1}{n-1} \Big/ \binom{\theta}{n}$$

$$= \underline{\hspace{4cm}}$$

$$\mathbb{E}(X_{(n)}) = \sum_{k=n}^{\theta} kP(X_{(n)} = k) = n \cdot \frac{\theta + 1}{n+1}$$

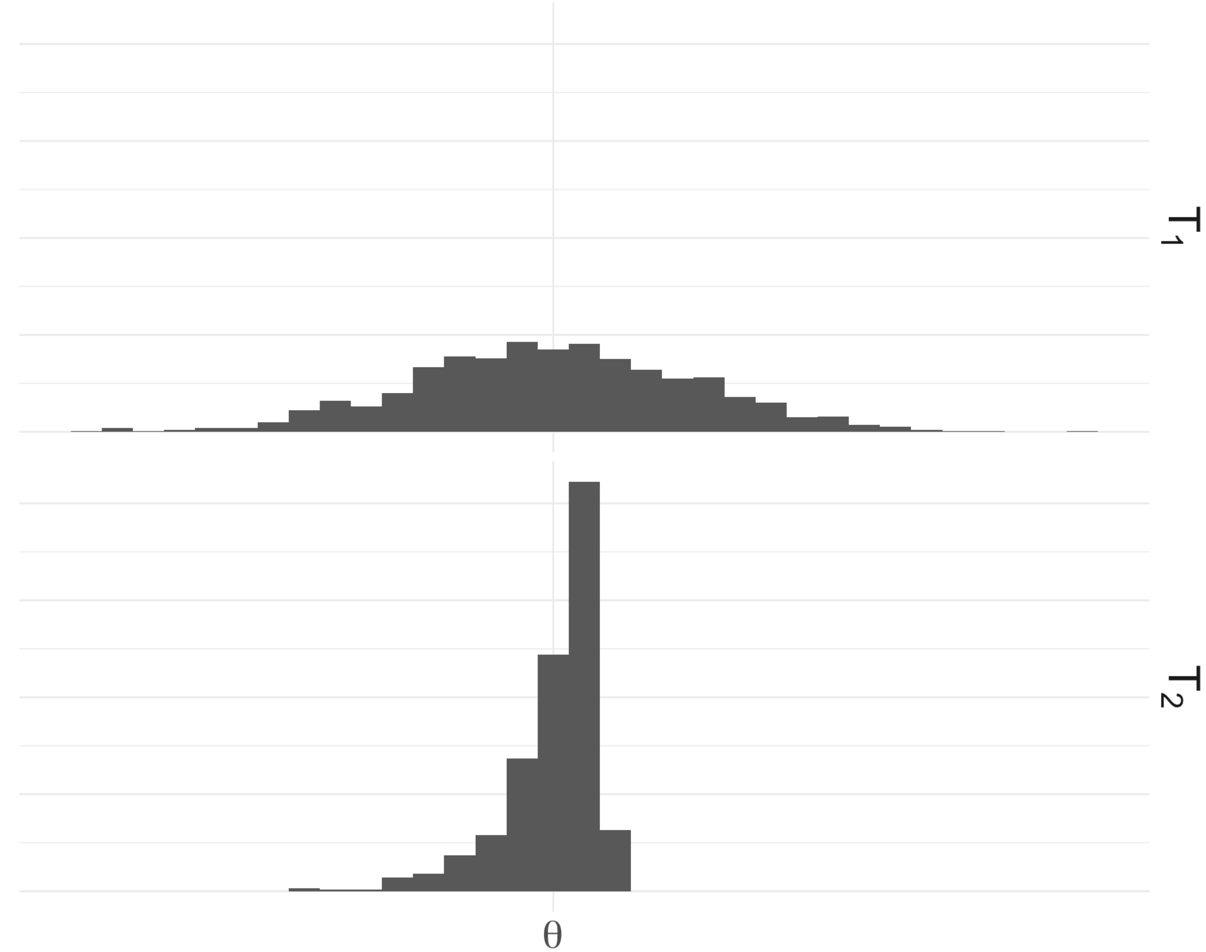
$$\implies T_2 = \frac{n+1}{n} X_{(n)} - 1$$

is an unbiased estimator for θ.

Which is *better?*

```r
m <- 1000
n <- 25
theta <- 1000
t1 <- numeric(m)
t2 <- numeric(m)
for (i in seq(m)) {
  x <- sample(seq(theta), n, replace = FA
  t1[i] <- 2 * mean(x) - 1
  t2[i] <- (n + 1) * max(x) / n - 1
}
```

One way to compare two unbiased estimators is to compare their variances.



Efficiency of estimators

Comparing efficiencies of two unbiased estimators

See Section 20.2 from Dekking *et al.* for the variances of T_1 and T_2 from the German tank example.

Let T_1 and T_2 be two unbiased estimators for the same parameter θ. Then estimator T_2 is called **more efficient** than estimator T_2 when

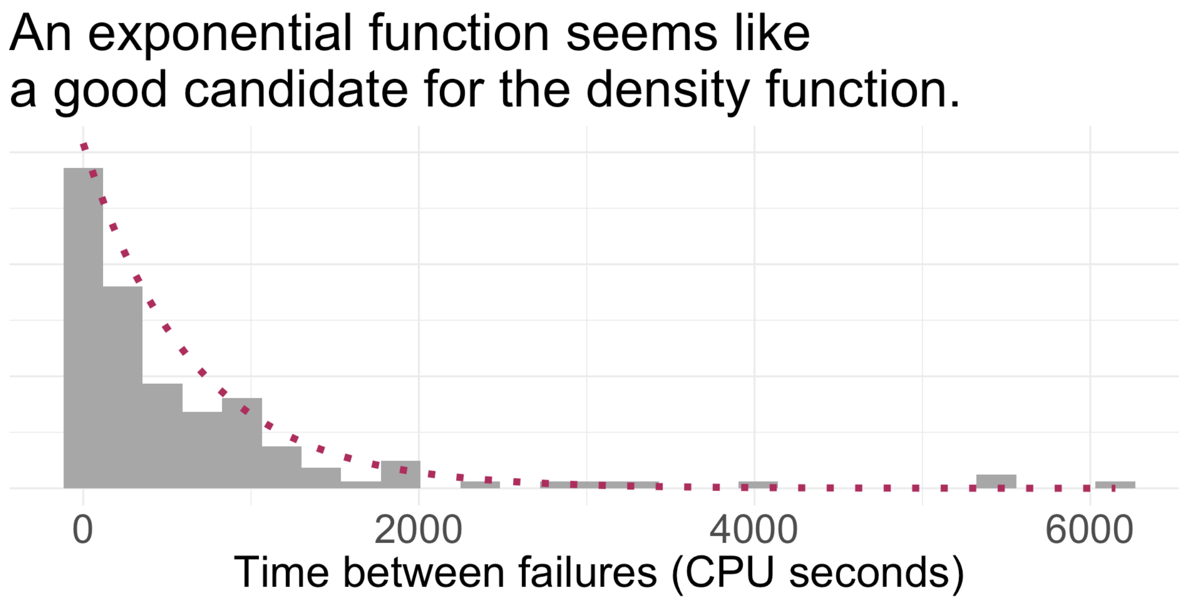
$$\mathrm{Var}(T_2) < \mathrm{Var}(T_1),$$

irrespective of the value of θ.

The **relative efficiency** of T_2 with respect to T_1 is

$$\frac{\mathrm{Var}(T_1)}{\mathrm{Var}(T_2)}.$$

Example: Software reliability data



See Section 15.3 of Dekking *et al.* for details.

- Assume the data is a random sample from $X \sim \mathrm{Exp}(\lambda)$.
- $T = \bar{X}_n$ is an unbiased estimator for $1/\lambda$ where $\bar{X}_n$ is the sample mean.
- $\mathrm{Var}(T) = \frac{1}{n\lambda^2}$.

> How efficient is this estimator?

The Cramér-Rao inequality

> The lower bound provides the a bound on the *best efficiency* an unbiased can achieve.

Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from a continuous distribution with probability density function f_θ, where θ is the parameter of interest, and T is an unbiased estimator for θ.

Let X denote the population random variable whose probability density function is f_θ. Under certain smoothness conditions on the density f_θ, the **Cramér-Rao lower bound** states that

$$\mathrm{Var}(T) \geq \frac{1}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta\left(X\right)\right)^2\right]}$$

for all θ.

Example: Software reliability data

$$\mathrm{Var}(T) \geq \frac{1}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta\left(X\right)\right)^2\right]}$$

where $f_\theta(X) = \frac{1}{\theta}e^{-X/\theta}$.

Example: Software reliability data

$$\mathrm{Var}(T) \geq \frac{1}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right]}$$

where $f_\theta(X) = \frac{1}{\theta}e^{-X/\theta}$.

$$\implies \mathrm{Var}(T) \geq \frac{1}{n\lambda^2}$$

> The sample mean is the most efficient estimator for λ. When an unbiased estimator achieves the Cramér-Rao bound, we call the estimator a **minimum variance unbiased estimator**.

$$\log(f_\theta(X)) = \log\left(\frac{1}{\theta}\right) - \frac{X}{\theta} = -\log(\theta) - \frac{X}{\theta}$$

$$\begin{aligned}\left[\frac{\partial}{\partial\theta}\log(f_\theta(X))\right]^2 &= \left[-\frac{1}{\theta} + \frac{X}{\theta^2}\right]^2 \\ &= \frac{1}{\theta^2} - 2\frac{X}{\theta^3} + \frac{X^2}{\theta^4} \\ &= \lambda^2 - 2\lambda^3 X + \lambda^4 X^2 \quad (\theta = \frac{1}{\lambda})\end{aligned}$$

Take the expectation...

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log(f_\theta(X))\right]^2 = \lambda^2 - \lambda^4\mathbb{E}(X^2)$$

Use $\mathrm{Var}(X) = 1/\lambda^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$...

$$\ldots = \lambda^2$$

Exercises

Suppose you $U_1, U_2, \ldots, U_n \overset{i.i.d}{\sim} \mathrm{Unif}(0, \theta)$. Let T_1 be the method of moment estimator for θ and T_2 be the maximum $U_{(n)}$.

1. Which of the two estimators are more efficient?
2. Construct an unbiased estimator based on T_2 and call it T_3. Is T_3 more efficient than T_1? How about T_2?

Summary

- Method of moments is a simple way to construct estimators based on sample moments.
- For unbiased estimators, we can compare their **efficiency** using their variances.
- For unbiased estimators, the Cramér-Rao lower bound provides a theoretical minimum variance they can achieve based on the model distribution.

Example: Coffee shop

Suppose the daily count of coffees sold at Michael's coffee shop, Y, follows a Poisson distribution. The daily counts from the past week, $Y_1, Y_2, \ldots, Y_7$, are shown below. Assume they form a random sample.

2 4 2 5 6 0 3

We can estimate $p_0 = P(Y = 0)$ with

1. $T_1 = \frac{\sum_{i=1}^{7} \mathbb{I}(Y_i = 0)}{7}$
2. $T_2 = e^{-\bar{Y}_7}$

Example: Unbiased T_1 vs biased T_2

Recall ...

$$\mathbb{E}(T_1) = p_0$$

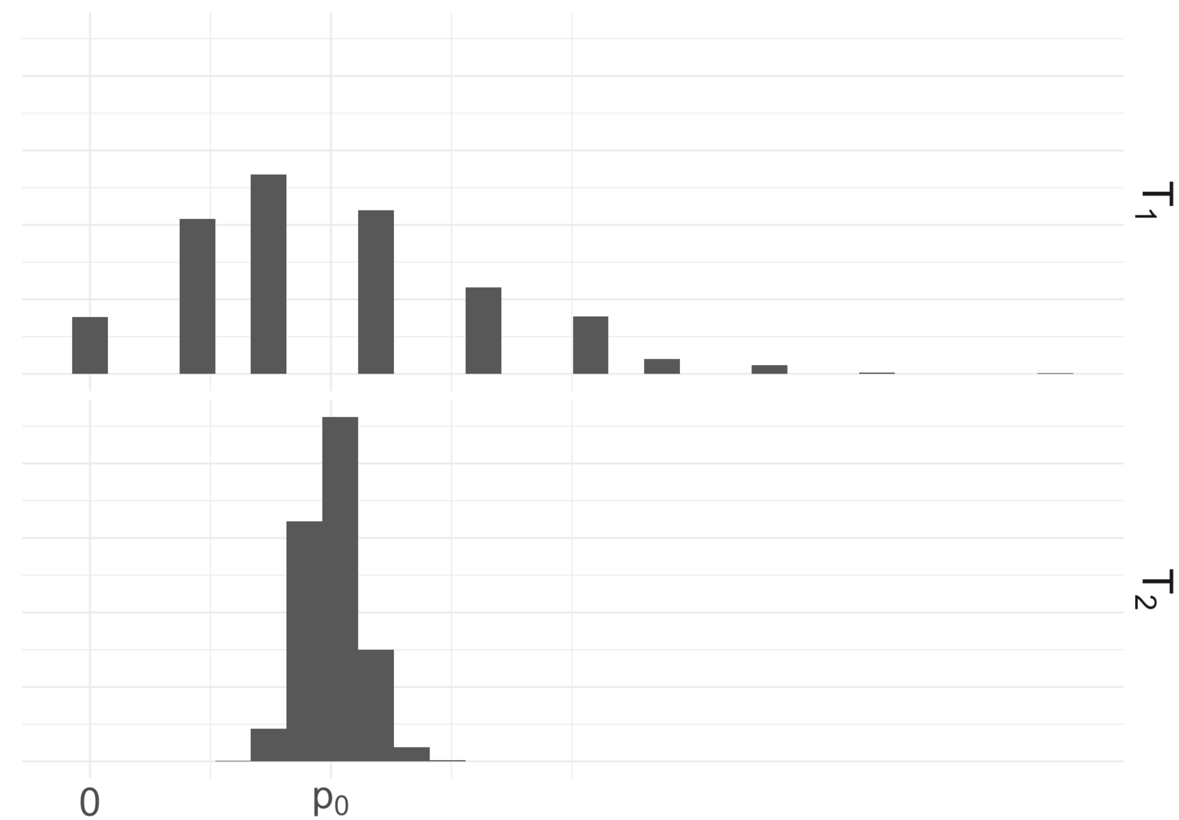
T_1 is unbiased.

$$\mathbb{E}(T_2) > p_0$$

T_2 is biased.

Example: Unbiased T_1 vs biased T_2

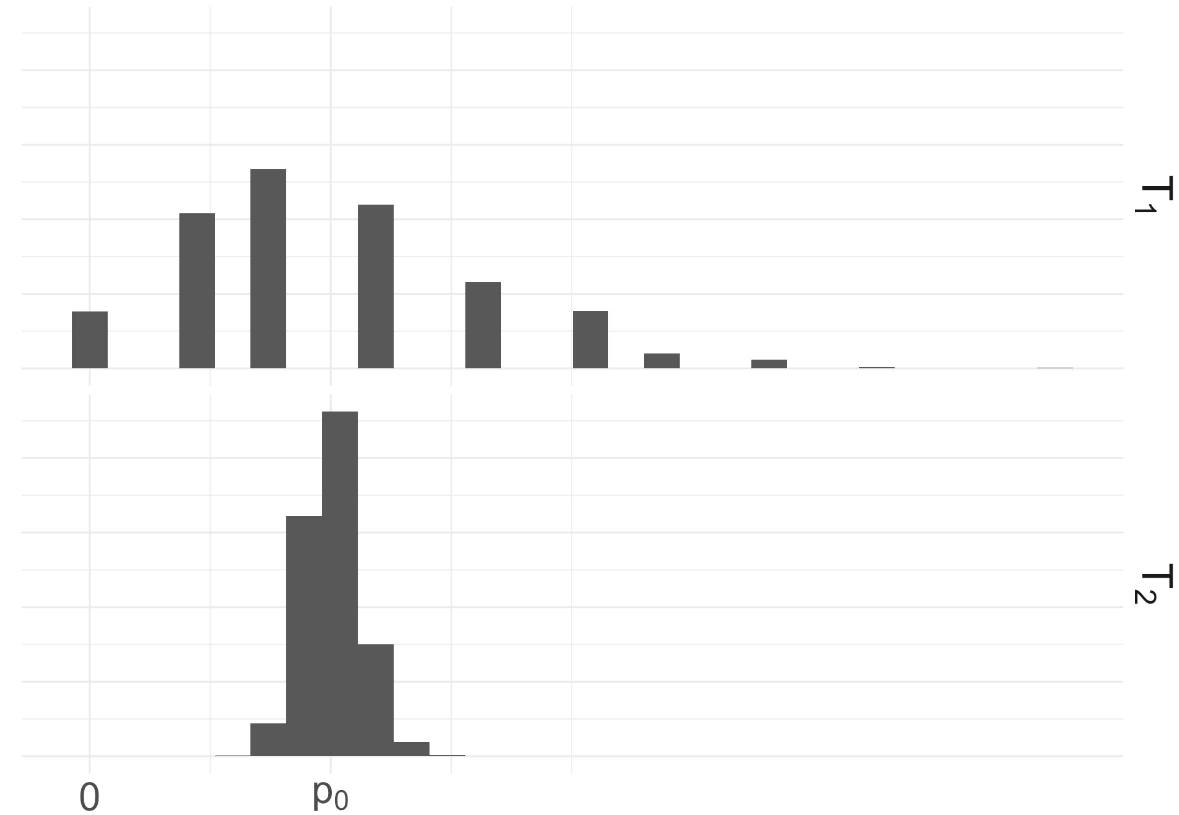
```r
m <- 1000
lambda <- 5
n <- 365 # for 1 year
t1 <- numeric(m)
t2 <- numeric(m)
for (i in seq(m)) {
  sales <- rpois(n, lambda)
  t1[i] <- mean(sales == 0)
  t2[i] <- exp(-mean(sales))
}
```



Which is more efficient?

Example: Unbiased T_1 vs biased T_2

```r
m <- 1000
lambda <- 5
n <- 365 # for 1 year
t1 <- numeric(m)
t2 <- numeric(m)
for (i in seq(m)) {
  sales <- rpois(n, lambda)
  t1[i] <- mean(sales == 0)
  t2[i] <- exp(-mean(sales))
}
```



Which is more efficient?

```r
var(t1)
```

```
[1] 1.812066e-05
```

```r
var(t2)
```

```
[1] 6.149703e-07
```

> $\text{Var}(T_1) > \text{Var}(T_2)$ but T_2 is biased.

Mean squared error

We can use the mean squared error to compare *efficiencies* of estimators.

Let T be an estimator for a parameter θ. The **mean squared error** of T is the defined as

$$\mathrm{MSE}(T) = \mathbb{E}\left[(T-\theta)^2\right].$$

Why MSE?

- It’s the expected value of squared distance of the estimator from the parameter, or the error.
- The smaller the mean squared error, the more efficient an estimator.

Why MSE?

- It's the expected value of squared distance of the estimator from the parameter, or the error.
- The smaller the mean squared error, the more efficient an estimator.
- It accounts for both the bias and the variance of an estimator.

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= \mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2\right] \\
&= \mathrm{Var}\left(\hat{\theta}-\theta\right) + \left[\mathbb{E}\left(\hat{\theta}-\theta\right)\right]^2 \\
&= \mathrm{Var}\left(\hat{\theta}\right) + \left[\mathrm{bias}\left(\hat{\theta}\right)\right]^2
\end{aligned}$$

> $$\implies \mathrm{MSE}(\hat{\theta}) = \mathrm{Var}\left(\hat{\theta}\right) \iff \mathrm{bias}\left(\hat{\theta}\right) = 0$$

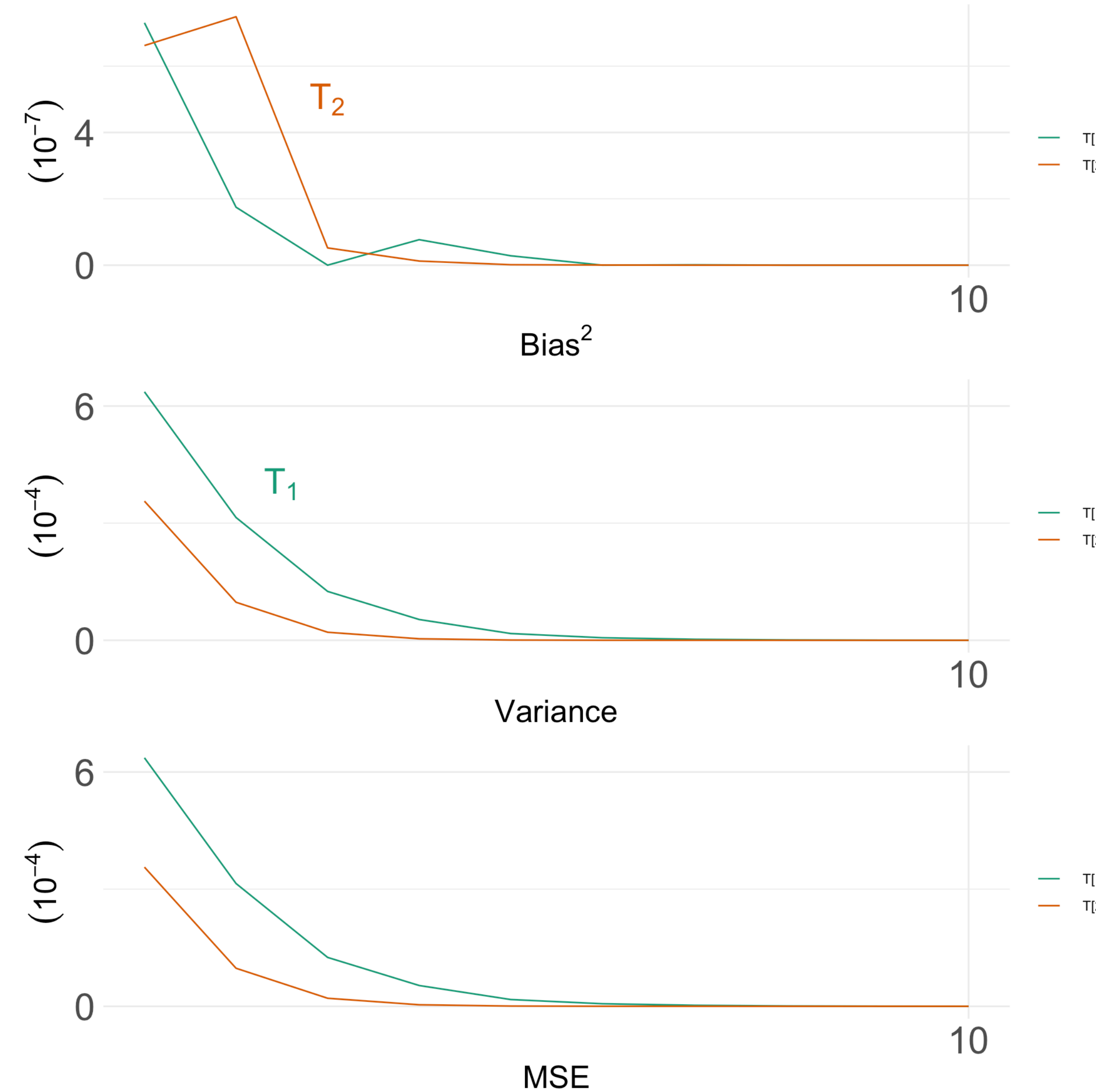
> Given two estimators, T_1 and T_2, for a parameter θ, we can say T_1 is a **more efficient estimator** when
>
> $$\mathrm{MSE}(T_1) < \mathrm{MSE}(T_2)$$
>
> irrespective of the value of θ.

Example: Is T_2 more efficient?

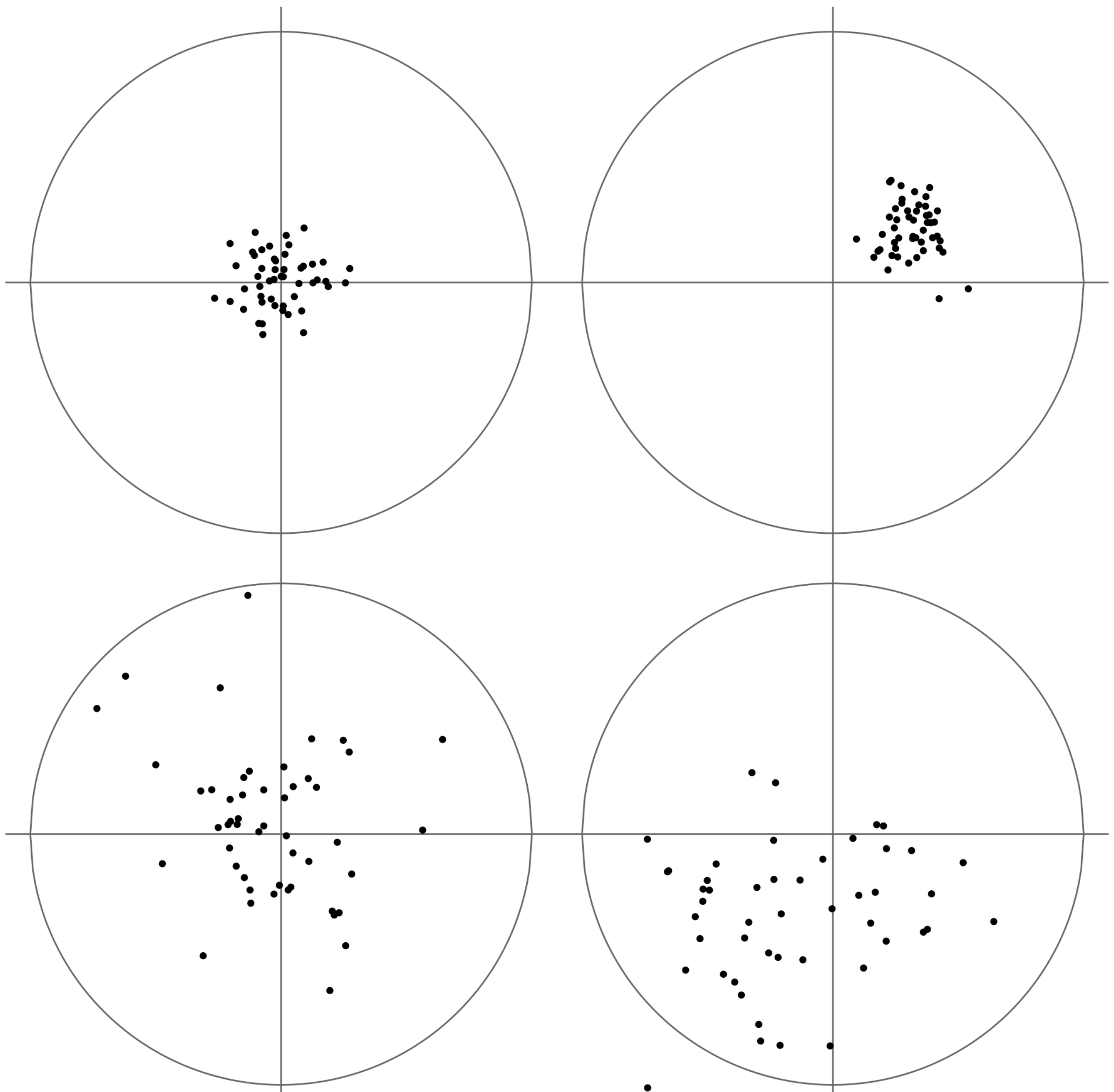
```r
lambdas <- seq(10)
bias_t1 <- numeric(10)
bias_t2 <- numeric(10)
var_t1 <- numeric(10)
var_t2 <- numeric(10)

m <- 1000
n <- 365 # for 1 year

for (lambda in lambdas) {
  t1 <- numeric(m)
  t2 <- numeric(m)
  for (i in seq(m)) {
    sales <- rpois(n, lambda)
    t1[i] <- mean(sales == 0)
    t2[i] <- exp(-mean(sales))
  }
  bias_t1[lambda] <- mean(t1) - exp(-lambda)
  bias_t2[lambda] <- mean(t2) - exp(-lambda)
  var_t1[lambda] <- var(t1)
  var_t2[lambda] <- var(t2)
}
```



Bias-variance tradeoff



We may need to choose between an unbiased estimator and a more “efficient” estimator based on the MSE. Make a conscious decision based on the tradeoff.

Exercises

Suppose $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} F$. For each of the estimators $\hat{\theta}$, compute the MSE.

1. F is $\text{Unif}(0, \theta)$ and $\hat{\theta} = X_{(n)}$.
2. F is $\text{Binom}(10, \theta)$ and $\hat{\theta} = \bar{X}_n/m$.
3. F is $\mathcal{N}(\theta, 1)$ and $\hat{\theta} = \bar{X}_n$.

Summary

- Mean squared error accounts for both bias and variance.
- Unbiased estimators are not always the most efficient based on their MSEs.
- Sometimes we need to make the conscious tradeoff between bias and variance.