Lecture 2. Estimation, bias, and mean squared error

Estimators

- Suppose that $X_1, \dots, X_n$ are iid, each with pdf/pmf $f_X(x \mid \theta)$, θ unknown.
- We aim to estimate θ by a **statistic**, ie by a function T of the data.
- If $\mathbf{X} = \mathbf{x} = (x_1, \dots, x_n)$ then our estimate is $\hat{\theta} = T(\mathbf{x})$ (does not involve θ).
- Then $T(\mathbf{X})$ is our **estimator** of θ, and is a rv since it inherits random fluctuations from those of $\mathbf{X}$.
- The distribution of $T = T(\mathbf{X})$ is called its **sampling distribution**.

Example

Let $X_1, \dots, X_n$ be iid $N(\mu, 1)$.

A possible estimator for μ is $T(\mathbf{X}) = \frac{1}{n}\sum X_i$.

For any particular observed sample $\mathbf{x}$, our estimate is $T(\mathbf{x}) = \frac{1}{n}\sum x_i$.

We have $T(\mathbf{X}) \sim N(\mu, 1/n)$. $\square$

Bias

If $\hat{\theta} = T(\mathbf{X})$ is an estimator of θ, then the *bias* of $\hat{\theta}$ is the difference between its expectation and the 'true' value: i.e.

$$\text{bias}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}) - \theta.$$

An estimator $T(\mathbf{X})$ is **unbiased** for θ if $\mathbb{E}_\theta T(\mathbf{X}) = \theta$ for all θ, otherwise it is **biased**.

In the above example, $\mathbb{E}_\mu(T) = \mu$ so T is unbiased for μ.

[Notation note: when a parameter subscript is used with an expectation or variance, it refers to the parameter that is being conditioned on. i.e. the expectation or variance will be a function of the subscript]

Mean squared error

Recall that an estimator T is a function of the data, and hence is a random quantity. Roughly, we prefer estimators whose sampling distributions "cluster more closely" around the true value of θ, whatever that value might be.

Definition 2.1

The **mean squared error** (mse) of an estimator $\hat{\theta}$ is $\mathbb{E}_\theta\left[(\hat{\theta} - \theta)^2\right]$.

For an unbiased estimator, the mse is just the variance.

In general

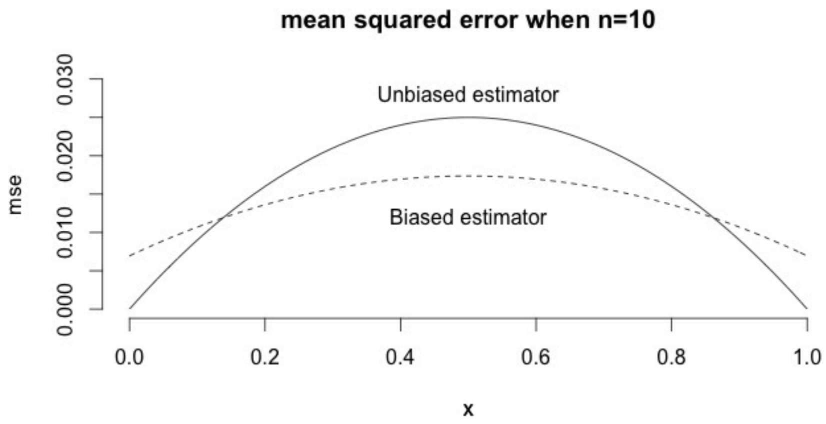
$$\begin{aligned}
\mathbb{E}_\theta\left[(\hat{\theta} - \theta)^2\right] &= \mathbb{E}_\theta\left[(\hat{\theta} - \mathbb{E}_\theta\hat{\theta} + \mathbb{E}_\theta\hat{\theta} - \theta)^2\right] \\
&= \mathbb{E}_\theta\left[(\hat{\theta} - \mathbb{E}_\theta\hat{\theta})^2\right] + \left[\mathbb{E}_\theta(\hat{\theta}) - \theta\right]^2 + 2\left[\mathbb{E}_\theta(\hat{\theta}) - \theta\right]\mathbb{E}_\theta\left[\hat{\theta} - \mathbb{E}_\theta\hat{\theta}\right] \\
&= \text{var}_\theta(\hat{\theta}) + \text{bias}^2(\hat{\theta}),
\end{aligned}$$

where $\text{bias}(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}) - \theta$.

[NB: sometimes it can be preferable to have a biased estimator with a low variance - this is sometimes known as the 'bias-variance tradeoff'.]

Example: Alternative estimators for Binomial mean

- Suppose $X \sim \text{Binomial}(n, \theta)$, and we want to estimate θ.
- The standard estimator is $T_U = X/n$, which is Unbiassed since $\mathbb{E}_\theta(T_U) = n\theta/n = \theta$.
- T_U has variance $var_\theta(T_U) = var_\theta(X)/n^2 = \theta(1-\theta)/n$.
- Consider an alternative estimator $T_B = \frac{X+1}{n+2} = w\frac{X}{n} + (1-w)\frac{1}{2}$, where $w = n/(n+2)$.
- (Note: T_B is a weighted average of X/n and $\frac{1}{2}$.)
- e.g. if X is 8 successes out of 10 trials, we would estimate the underlying success probability as $T(8) = 9/12 = 0.75$, rather than 0.8.
- Then $\mathbb{E}_\theta(T_B) - \theta = \frac{n\theta+1}{n+2} - \theta = (1-w)\left(\frac{1}{2} - \theta\right)$, and so it is biased.
- $var_\theta(T_B) = \frac{var_\theta(X)}{(n+2)^2} = w^2\theta(1-\theta)/n$.
- Now $mse(T_U) = var_\theta(T_U) + \text{bias}^2(T_U) = \theta(1-\theta)/n$.
- $mse(T_B) = var_\theta(T_B) + \text{bias}^2(T_B) = w^2\theta(1-\theta)/n + (1-w)^2\left(\frac{1}{2} - \theta\right)^2$



So the biased estimator has smaller MSE in much of the range of θ

T_B may be preferable if we do not think θ is near 0 or 1.

So our *prior judgement* about θ might affect our choice of estimator.

Will see more of this when we come to Bayesian methods,.

Why unbiasedness is not necessarily so great

Suppose $X \sim \text{Poisson}(\lambda)$, and for some reason (which escapes me for the moment), you want to estimate $\theta = [\mathbb{P}(X=0)]^2 = e^{-2\lambda}$.

Then any unbiassed estimator $T(X)$ must satisfy $\mathbb{E}_\theta(T(X)) = \theta$, or equivalently

$$\mathbb{E}_\lambda(T(X)) = e^{-\lambda}\sum_{x=0}^{\infty} T(x)\frac{\lambda^x}{x!} = e^{-2\lambda}.$$

The only function T that can satisfy this equation is $T(X) = (-1)^X$ [coefficients of polynomial must match].

Thus the only unbiassed estimator estimates $e^{-2\lambda}$ to be 1 if X is even, -1 if X is odd.

This is not sensible.