1. The empirical distribution

Definition 1. *If $y_1, \dots, y_n \in \mathbb{R}$, then the cumulative distribution function $\widehat{F}_y$ defined by*

$$\widehat{F}_y(x) := \frac{1}{n}\sum_{j=1}^{n} I(Y_j \le x) = \frac{\#\{j : y_j \le x\}}{n}.$$

is called the empirical distribution function *associated to $y_1, \dots, y_n$.*

If $Y_1, \dots, Y_n \overset{\text{IID}}{\sim} F$, we may unambiguously set

$$\widehat{F}_n(x) := \widehat{F}_Y(x),$$

since this random variable depends only on F, n, and x.

Theorem 2. *Let $x \in \mathbb{R}$ and let $Y_1, \dots, Y_n \overset{\text{IID}}{\sim} F$. Then:*

(1) $\mathrm{E}\,\widehat{F}_Y(x) = F(x)$
(2) $\mathrm{Var}\,\widehat{F}_Y(x) = \frac{1}{n}F(x)(1 - F(x))$

Proof. (1) follows from the identities

$$\mathrm{E}\,I(Y_j \le x) = \mathrm{Prob}(Y_j \le x) = F(x).$$

To prove (2), we compute:

$$\begin{aligned}
\mathrm{Var}\,\widehat{F}_Y(x) &= \frac{1}{n^2}\sum_{j=1}^{n} \mathrm{Var}\,I(Y_j \le x) \\
&= \frac{1}{n^2}\sum_{j=1}^{n} \big(\mathrm{E}\,I(Y_j \le x)^2 - (\mathrm{E}\,I(Y_j \le x))^2\big) \\
&= \frac{1}{n^2}\sum_{j=1}^{n} \big(\mathrm{E}\,I(Y_j \le x) - (\mathrm{E}\,I(Y_j \le x))^2\big) && (\text{as } I^2 = I) \\
&= \frac{1}{n^2} n(F(x) - F(x)^2) \\
&= \frac{1}{n}F(x)(1 - F(x))
\end{aligned}$$

□

2. Substitution estimators

Let $\mathscr{P}$ be the set of all distributions on $\mathbb{R}$. For a statistical functional $\theta : \mathscr{P} \to \mathbb{R}$, define $\widehat{\theta} : \mathscr{X} \to \mathbb{R}$ by the rule

$$\widehat{\theta}(y_1, \dots, y_n) = \theta(\widehat{F}_y).$$

Definition 3. *If $Y_1, \dots, Y_n \overset{\text{IID}}{\sim} F$, then $\widehat{\theta}(Y_1, \dots, Y_n)$ is called the* substitution estimator *of θ associated to $Y_1, \dots, Y_n$.*

Formally, we get $\widehat{\theta}(Y_1, \dots, Y_n)$ by substituting $\widehat{F}_n$ for F in $\theta(F)$:

$$\widehat{\theta}(Y_1, \dots, Y_n) = \theta(\widehat{F}_n).$$

Making sense of the expression $\theta(\widehat{F}_n)$ leads to Definition 3.

Let

$$\mu : \mathscr{P} \to \mathbb{R}, \quad \mu(F) = \int_{-\infty}^{\infty} x \, dF(x)$$

be the mean functional. We compute the *empirical mean*, $\widehat{\mu}(Y_1, \dots, Y_n)$, associated to $Y_1, \dots, Y_n \overset{\text{IID}}{\sim} F$:

$$\begin{aligned}
\widehat{\mu}(Y_1, \dots, Y_n) &= \mu(\widehat{F}_n) \\
&= \int_{-\infty}^{\infty} x \, d\widehat{F}_n(x) \\
&= \frac{1}{n} \int_{-\infty}^{\infty} x \, dI(Y_j \leq x) \\
&= \frac{1}{n} \sum_{j=1}^{n} Y_j \\
&= \overline{Y}
\end{aligned}$$

Thus, the empirical mean is just the sample mean.

Let

$$\sigma^2 : \mathscr{P} \to \mathbb{R}, \quad \mu(F) = \int_{-\infty}^{\infty} (x - \mu(F))^2 \, dF(x)$$

be the variance functional. We compute the *empirical variance*, $\widehat{\sigma^2}(Y_1, \dots, Y_n)$, associated to $Y_1, \dots, Y_n \overset{\text{IID}}{\sim} F$:

$$\begin{aligned}
\widehat{\sigma^2}(Y_1, \dots, Y_n) &= \sigma^2(\widehat{F}_n) \\
&= \int_{-\infty}^{\infty} (x - \mu(\widehat{F}_n))^2 \, d\widehat{F}_n(x) \\
&= \frac{1}{n} \int_{-\infty}^{\infty} (x - \overline{Y})^2 \, dI(Y_j \leq x) \\
&= \frac{1}{n} \sum_{j=1}^{n} (Y_j - \overline{Y})^2
\end{aligned}$$

The empirical variance $\widehat{\sigma^2}$ and the sample variance, S^2, are different:

$$\widehat{\sigma^2} = \frac{n-1}{n} S^2, \quad \text{where} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

Consider $T_b := bS^2$ as an estimate of σ^2.

$$\mathrm{MSE}(T_b, \sigma^2) = \operatorname{Var} T_b + \operatorname{Bias}(T_b, \sigma)^2$$

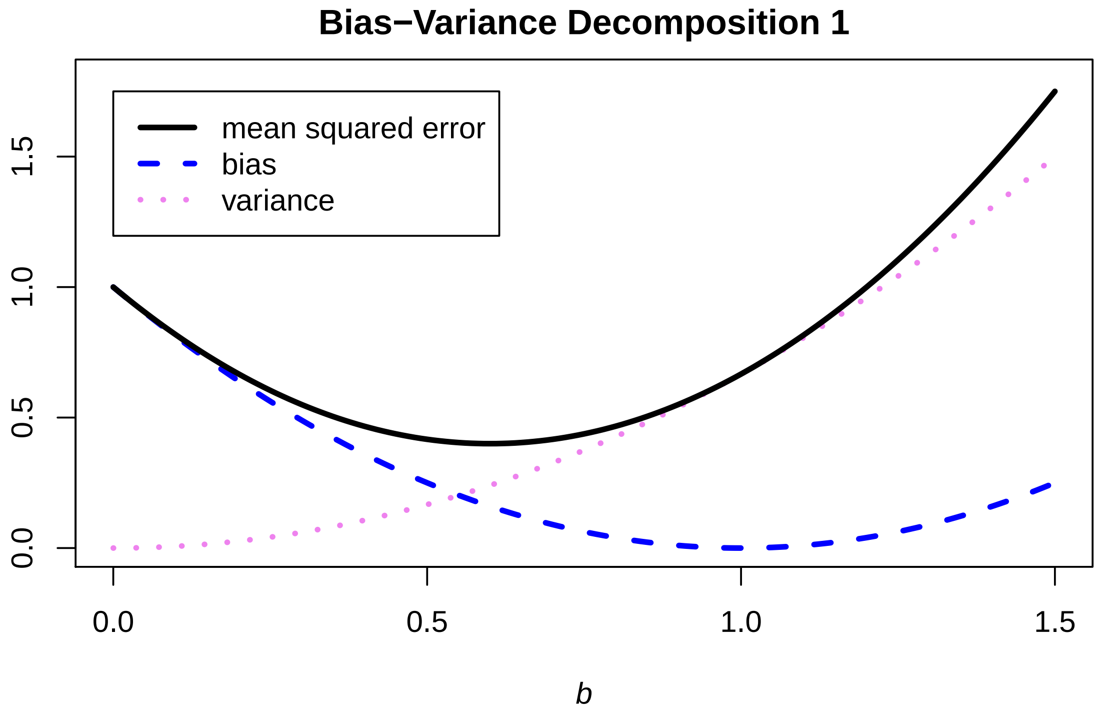

FIGURE 1. Plot of b vs $\operatorname{MSE}(T_b, \sigma^2)$, $\operatorname{Bias}(T_b, \sigma^2)^2$, and $\operatorname{Var} T_b$ for $n = 5$

$$\operatorname{Var} T_b = \operatorname{Var} bS^2 = b^2 \operatorname{Var} S^2 = b^2 \frac{2\sigma^4}{n-1} = \frac{2b^2}{n-1}\sigma^4$$

$$\operatorname{Bias}(T_b, \sigma^2) = \operatorname{E}[T_b] - \sigma^2 = \operatorname{E}[bS^2] - \sigma^2 = b\operatorname{E}[S^2] - \sigma^2 = b\sigma^2 - \sigma^2 = (b-1)\sigma^2$$

Therefore,

$$\operatorname{Bias}(T_b, \sigma^2)^2 = (b-1)^2\sigma^4.$$

$$\begin{aligned}\operatorname{MSE}(T_b, \sigma^4) = \frac{2b^2}{n-1}\sigma^4 + (b-1)^2\sigma^4 &= \left(\frac{2b^2}{n-1} + (b-1)^2\right)\sigma^4 \\ &= \frac{1}{n-1}\big((n+1)b^2 - 2(n-1)b + n - 1\big)\,\sigma^4\end{aligned}$$

As a function of b, $\operatorname{MSE}(T_b, \sigma^2)$ is minimized when $(n+1)b^2 - 2(n-1)b + n - 1$, i.e., at

$$b = \frac{n-1}{n+1}.$$

3. Density estimation

3.1. **A point estimate.** Let $\mathscr{P}$ be the set of all probability distributions that have smooth densities. Since a smooth density characterizes a distribution uniquely (why?), we can identify $\mathscr{P}$ with the set of all smooth probability density functions on $\mathbb{R}$:

$$\mathscr{P} = \left\{ f : \mathbb{R} \to [0, \infty) : f \text{ is smooth and } \int_{-\infty}^{\infty} f(x)dx = 1 \right\}.$$

Define a statistical functional θ on $\mathscr{P}$ by rule

$$\theta(f) = f(0).$$

Let $X_1, \ldots, X_n \overset{\text{IID}}{\sim} f$ and let $h > 0$. Then

$$\begin{aligned} p_h &:= \text{Prob}(|X_j| \le h/2) \\ &= \int_{-h/2}^{h/2} f(x)\,dx \\ &= \int_{-h/2}^{h/2} \left(f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(0)}{6}h^3 + O(h^4) \right) dx \\ &= f(0)h + \frac{f''(0)}{24}h^3 + O(h^5). \end{aligned}$$

Thus, for small h,

$$\mathrm{E}\left[\frac{I(|X_j| \le h/2)}{h} \right] \approx \theta(f).$$

In particular, for small h,

This motivates considering

$$\widehat{\theta}(X_1, \ldots, X_n) := \frac{1}{nh} \sum_{j=1}^{n} I(|X_j| \le h/2)$$

as an estimator of $\theta(f)$. Then

$$\mathrm{E}\,\theta(X_1, \ldots, X_n) = f(0) + \frac{f''(0)}{24}h^2 + O(h^3)$$

and

$$\text{Bias}(\widehat{\theta}, \theta)^2 = \frac{f''(0)^2}{576}h^4 + O(h^6).$$

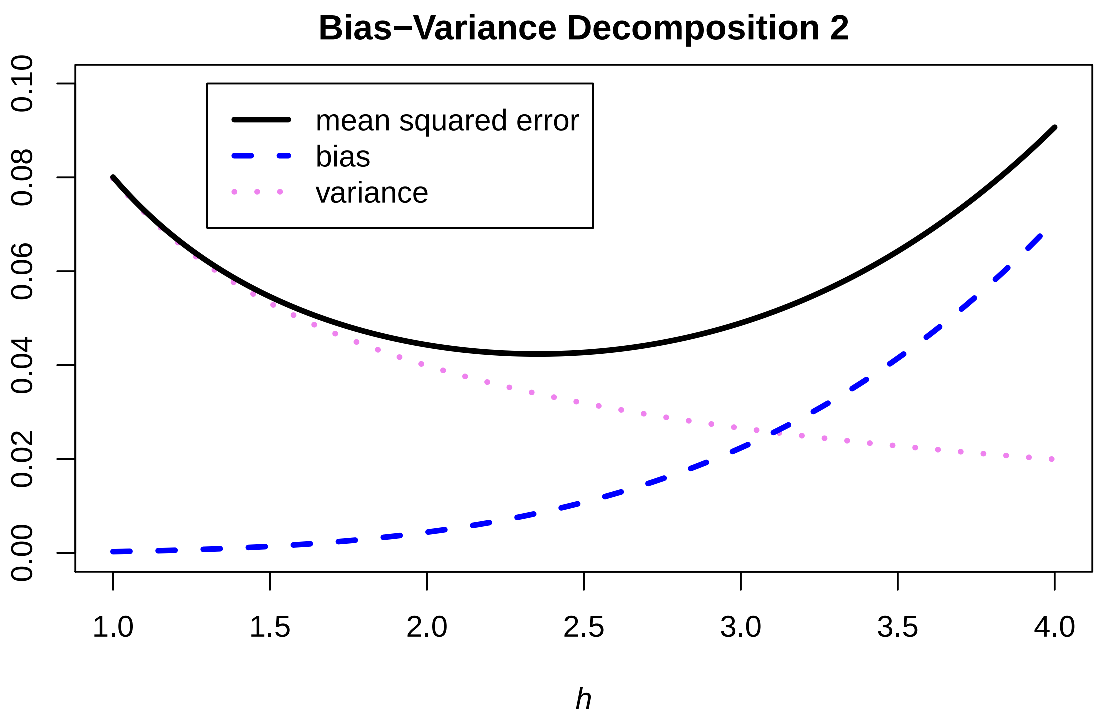

FIGURE 2. Plot of b vs $\mathrm{MSE}(\widehat{\theta}, \theta)$, $\mathrm{Bias}(\widehat{\theta}, \theta)^2$, and $\mathrm{Var}\, \widehat{\theta}$ for $n = 5$

As for the variance,

$$\begin{aligned}
\mathrm{Var}\, \widehat{\theta}(X_1, \ldots, X_n) &= \frac{1}{n^2 h^2} \sum_{j=1}^{n} \mathrm{Var}\, I(|X_j| \leq h/2) \\
&= \frac{1}{n^2 h^2} n p_h (1 - p_h) \\
&= \frac{1}{n h^2} \left(f(0)h - f(0)^2 h^2 + \frac{f''(0)}{24} h^3 - \frac{f(0) f''(0)}{12} h^4 + \frac{f''(0)^2}{576} h^6 \right) \\
&= \frac{1}{n} \left(\frac{f(0)}{h} - f(0)^2 + \frac{f''(0)}{24} h - \frac{f(0) f''(0)}{12} h^2 + O(h^4) \right)
\end{aligned}$$

3.2. **The histogram estimator.**

4. THE METHOD OF MAXIMUM LIKELIHOOD

$E[Y|X] = aX + b$

Suppose that

$$Y = f(X) + \varepsilon$$