The process of estimating a statistic is called estimation. The statistical method you are using in the estimation is called an estimator. (e.g. the sample mean, the mode, median)

Estimation is plagued by several problems.

1. Irregular data (i.e. outliers)
2. Scarcity – not having access to all the data.

Depending on the goal different statistics (i.e. estimators) will be appropriate.

To Emphasize the distribution's dependence on its parameter θ

$f(y|\theta) = \{ \theta e^{-\theta y}, \; y > 0; \; 0, \text{ otherwise}$

① Minimizing error.

Error Calculation $\{ MSE = \frac{1}{m} \sum (\bar{x} - \mu)^2$. Calculate the euclidean distance based on the sum of deviations between the estimations and the true mean.

$RMSE = (MSE)^{\frac{1}{2}}$

$= \sqrt{\frac{1}{m} \sum (\bar{x} - \mu)^2}$

m – is the number of trials

n – is the size of the sample used to calculate $\bar{x}$.

(*) Cost as a function of error is not symmetric.

② Highest chance of being right.

is the Maximum Likelihood Estimator (MLE)

Estimating Variance

Sample the (σ^2) V → Population Variance

→ Small Sample Variance

S_{n-1}^2

Method

Statistic to Calculate

Intent/Desire

☐ Sample Variance

$\sigma^2 \approx S^2$

$\Leftrightarrow S^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$

– Accurate for large samples

– Poor for small samples.

This is called a biased estimator.

Unfortunately the total error or mean deviation from the variance will not converge to zero.

☐ Unbiased Estimator of Variance.

The error has the expectation of decreasing to zero with more iterations

$S_{n-1}^2 = \frac{1}{(n-1)} \sum (x_i - \bar{x})^2$

What are degrees of freedom

"The number of dimensions (pieces of information) in the data which are free to vary when estimating statistical parameters.

Degrees of Freedom = num. of dimensions – num. required relations among dimensions (e.g. num. param estimated).

For example, in a 1-sample t-test, one degree of freedom is spent estimating μ – the mean.

$\vec{v} = [10, 2, 9, x] \longrightarrow \mu = \frac{1}{\dim(\vec{v})} \sum_i x_i$

When you perform regression, a parameter is estimated for every term in the model, each consumes a degree of freedom.

Including too many terms in a model reduces the number of degrees of freedom available to estimate the parameters variability.

Data allows for more degrees of freedom, which can be used to calculate the error term, p-value, and F-value.

$\boxed{\text{Bias Function}}$ Alternative goodness of an Estimator – Error MSE, Relative Efficiency, Variance

is the difference between the estimator's expected [illegible] expected value and the true value of the parameter being estimated.

Sample Mean
$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$
Sample Variance
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

- Not a pejorative term — a stats property
- Bias can be measured relative to the median, rather than mean expected value.

Given sample $Y_1, \dots, Y_n$, with common density of $f(y; \theta)$, we want to find $\hat{\theta} = g(Y_1, \dots, Y_n)$ a suitable estimator for the real-valued function of the random sample g.

Bias,

$$b(\hat{\theta}) := E[\hat{\theta}] - \theta$$

, $\hat{\theta}$ is unbiased if $b(\hat{\theta}) = 0$

Mean-Square Error (Measures the quality of an estimator)

$$MSE(\hat{\theta}) := E_{\theta}\left[(\hat{\theta} - \theta)^2\right]$$

Bias

Mean Squared Error

INTERPRETATION #2 — VIA INTUIT.

Proof of $MSE(\hat{\theta}) = Var_{\theta}(\hat{\theta}) + Bias(\hat{\theta}, \theta)^2$.

$MSE(\hat{\theta}) = E_{\theta}[(\hat{\theta} - \theta)^2]$

$= E_{\theta}[(\hat{\theta} - E_{\theta}[\hat{\theta}] + E_{\theta}[\hat{\theta}] - \theta)^2]$, adding zero

$= E_{\theta}[$
$(\hat{\theta} - E_{\theta}[\hat{\theta}])^2$
$+ 2(\hat{\theta} - E_{\theta}[\hat{\theta}])(E_{\theta}[\hat{\theta}] - \theta)$
$+ (E_{\theta}[\hat{\theta}] - \theta)^2]$, by binomial exp.

$= E_{\theta}[(\hat{\theta} - E_{\theta}[\hat{\theta}])^2]$
$+ E_{\theta}[2(\hat{\theta} - E_{\theta}[\hat{\theta}])(E_{\theta}[\hat{\theta}] - \theta)]$
$+ E_{\theta}[(E_{\theta}[\hat{\theta}] - \theta)^2]$, by expected additivity

$= E_{\theta}[(\hat{\theta} - E_{\theta}[\hat{\theta}])^2]$
$+ 2(E_{\theta}[\hat{\theta}] - \theta)E_{\theta}[(\hat{\theta} - E_{\theta}[\hat{\theta}])]$
$+ (E_{\theta}[\hat{\theta}] - \theta)^2$, given that const $= E_{\theta}[\text{const}] = \dots$

$= E_{\theta}[(\hat{\theta} - E_{\theta}[\hat{\theta}])^2]$
$+ 2(E_{\theta}[\hat{\theta}] - \theta)(E_{\theta}[\hat{\theta}] - E_{\theta}[\hat{\theta}])$
$+ (E_{\theta}[\hat{\theta}] - \theta)^2$, provided $E_{\theta}[\hat{\theta}] = \text{const}$.

$= E_{\theta}[(\hat{\theta} - E_{\theta}[\hat{\theta}])^2] + (E_{\theta}[\hat{\theta}] - \theta)^2$

$= Var_{\theta}(\hat{\theta}) + Bias_{\theta}(\hat{\theta}, \theta)^2$.

On the other hand the Mean Standard Error can be thought of an alternative function: one which approximates the estimated value. It is desirable to use the estimator with the least variance. Theoretically for the stability — this applies to unbiased estimators. *This may hold for asymptotically unbiased functions.

If the task is to choose between biased estimators, then a criteria for selection is either's asymptotic behaviour — and if that asymptotic return is unbiased.

$\sigma_{\hat{\theta}} = \sqrt{Var_{\theta}(\hat{\theta})}$ } The Standard Error

Example of Bias, Variance & MSE

Given the ~~random sample~~ $Y_1, \dots, Y_n$, with the distribution of,

$$f(y|\theta) = \begin{cases} 3\theta^3 y^{-4}, & \theta \le y, \\ 0, & \text{otherwise.} \end{cases}$$

Given the Estimator,

$$\hat{\theta} = \min\{Y_1, \dots, Y_n\}$$

Bias can be calculated,

$B_{\theta}(\hat{\theta}) = E_{\theta}[\hat{\theta}] - \theta$.

Determine $E_{\theta}[\hat{\theta}]$,

$P(\hat{\theta} > x) = P(\min\{Y_1, \dots, Y_n\} > x)$
$= P(Y_1 > x, Y_2 > x, \dots, Y_n > x)$
$= [P(Y_1 > x)]^n$. Since $\{Y_1, \dots, Y_n\}$ are random.

$\boxed{\text{That is IID — independent \& identically distributed}}$

Given that $P(Y_i > x) = \int_x^{\infty} f(y|\theta)\,dy$,

$P(Y_1 > x) = \int_x^{\infty} 3\theta^3 y^{-4}\,dy$
$= \theta^3 x^{-3}$, when $x \ge \theta$.

Given $P(\hat{\theta} > x) = [P(Y_1 > x)]^n$,

$$P(\hat{\theta} > x) = [\theta^3 x^{-3}]^n$$
$$= \theta^{3n} x^{-3n}, \text{ when } x \geq \theta.$$

The Cumulative Distribution Function (CDF) is,

$$F(x) = P(\hat{\theta} \leq x)$$
$$= 1 - P(\hat{\theta} > x)$$
$$= 1 - \theta^{3n} x^{-3n}, \quad x \geq \theta$$

$$F(x) = 0, \quad x < \theta$$

The density function (PDF) is,

$$f(x) = F'(x) = \begin{cases} 3n\theta^{3n} x^{-(3n+1)}, & x \geq \theta \\ 0, & x < \theta \end{cases}$$

Therefore $E_\theta(\hat{\theta})$ is,

$$E_\theta(\hat{\theta}) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$
$$= \int_{\theta}^{\infty} x \cdot 3n\theta^{3n} x^{-(3n+1)}\,dx$$
$$= 3n\theta^{3n} \int_{\theta}^{\infty} x^{-3n}\,dx, \quad x \text{ factor removes exponent.}$$
$$= 3n\theta^{3n} \cdot \frac{\theta^{-3n+1}}{3n-1}$$
$$= \frac{3n}{3n-1} \cdot \theta.$$

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$
$$= \frac{3n}{3n-1}\theta - \theta$$
$$= \frac{\theta}{3n-1}.$$

Variance $(\hat{\theta})$,

$$Var(\hat{\theta}) = MSE(\hat{\theta}) - B(\hat{\theta})^2$$
$$= \frac{2\theta^2}{(3n-1)(3n-2)} - \left[\frac{\theta}{3n-1}\right]^2$$
$$= \frac{\theta^2}{(3n-1)}\left[\frac{2}{3n-2} - \frac{1}{3n-1}\right]$$
$$= \frac{3n\theta^2}{(3n-1)^2(3n-2)}.$$

Standard Error,

$$\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$$
$$= \frac{\theta}{(3n-1)}\sqrt{\frac{3n}{(3n-2)}}.$$

The $MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2$,

$$MSE(\hat{\theta}) = \int_{-\infty}^{\infty} (x-\theta)^2 f(x)\,dx$$
$$= \int_{\theta}^{\infty} (x-\theta)^2 \cdot 3n\theta^{3n} x^{-3n-1}\,dx$$
$$= 3n\theta^{3n}\left[\int_{\theta}^{\infty} x^{1-3n}\,dx - 2\theta\int_{\theta}^{\infty} x^{-3n}\,dx + \theta^2\int_{\theta}^{\infty} x^{-3n-1}\,dx\right]$$
$$= 3n\theta^{3n}\left[\frac{\theta^{2-3n}}{3n-2} - 2\theta\frac{\theta^{1-3n}}{3n-1} + \theta^2\frac{\theta^{-3n}}{3n}\right]$$
$$= 3n\theta^2\left[\frac{1}{3n-2} - \frac{2}{3n-1} + \frac{1}{3n}\right]$$
$$= \theta^2\left[\frac{(3n-1)(3n-2) - 9n(n-1)}{(3n-1)(3n-2)}\right]$$
$$= \frac{2\theta^2}{(3n-1)(3n-2)}$$

Asymptotically Unbiased Estimators

$$\lim_{n\to\infty} B(\hat{\theta}) = \lim_{n\to\infty} \frac{\theta}{3n-1}$$
$$= 0.$$

Therefore $\hat{\theta}$ is asymptotically unbiased.
This is the first quality to check when both estimators are biased.

Comparing Unbiased Estimators

Unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ can be compared by their efficiency

$$Eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$$

The estimator with the smaller variance is preferred.

Maximum Likelihood Estimation (MLE)
(Estimation method)

A method for estimating the parameters of a statistical model, given some observations.

These model parameter estimates are calculated by finding the parameters which maximise the

— likelihood function given the observations.

Articulating The Inaccuracy of Sampling a Distribution

When sampling a distribution the sample could contain any of the values in that distribution.

Due to this randomness, the summary statistics you generate based on that sample are dependent on that randomness. These summary statistics will therefore describe in part (i.e.) the element of this randomness (that is, sampling error).

Various Estimators can be used to calculate sample summary statistics of the distribution — but due to the randomness/entropy of the act of sampling these statistics aren't rigorous. The degree of their rigour can be approximated from other summary statistics (e.g. standard error, confidence intervals); these statistics are produced from the analysis of the same statistic applied to a an approximate distribution.

" i.e. a simulation "

Problem: how accurate/rigorous is estimation when there is sampling error?

① Comparative tools like,
- MSE/RMSE
- Bias
- Relative Efficiency

→ these made the estimate as if it were the correct one.

Do not account for Sampling Error or Bias!

Can be used to analyse/quantify the margin and difference between the real distribution's summary stat and the estimator function and hypothetical distrib, or the real summary stat and the estimations of the hypothetical distributions.

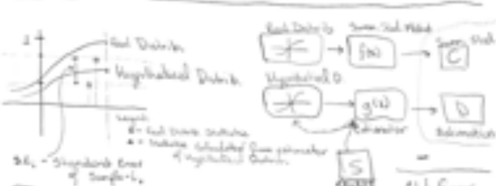
② Can also constrain the estimation or estimate via

- Confidence Intervals (CI): a range that includes a fraction of the sampling distribution. For example, 90% confidence interval is 5th percentile – 95th percentile.

Remember — Standard Error != S.D.

SE is the variability of error introduced by the estimator and the hypothetical distribution. This variability decreases → as the sample grows → thus is a snapshot of a refinement process.

SD calculates the variability in the data, ie the distribution which in reference to point in time is static.



* Hypothetical Distrib is formed from estimator/estimates from the sample. That Distrib is then used with the same estimator to approximate the distrib of samples

---

SAMPLING BIAS

Sampling error introduced by the sample process.

↳ Surveys are affected by factors and the exclusion of those affect the results of the survey.

- self-selection - people excluding themselves / refusing to answer the question.

- measurement error - rounding / fallibility

(*) when estimating a quantity, should provide
I) Standard Error
and/or
II) Confidence Interval

Exponential Distribution

- the mean of an exponential distrib. is $1/\lambda$.

- the λ param of a distrib. $\Rightarrow \mu = 1/\lambda \rightarrow \lambda = \frac{1}{\mu}$.

Both are biased.

↳ Maximum Likelihood Estimator of λ in $e(\lambda)$:

$$L = 1/\bar{x} \text{ , } \bar{x} \text{ is the sample mean.}$$

↳ Median-Based Likelihood (robust to outliers)

$$L_m = \ln(2)/m \text{ , } \text{median} = \ln(2)/\lambda.$$

$\bar{x}$ is unbiased of μ.

L is biased of λ param.