Russell Gayder
2019-03-09

Unbiased Parameter Estimation

[Following Jaynes' book, sections 13.8 and 9, 17.2]

Estimate parameter α from data $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

using a function of the data - an "estimator" $\beta(\underline{x})$.

An "unbiased" estimate has the property $\langle \beta(\underline{x}) \rangle = \alpha$.

Given density $f(\underline{x}|\alpha)$,

$$\langle \beta(\underline{x}) \rangle = \int \beta(\underline{x})\, f(\underline{x}|\alpha)\, d\underline{x}$$

$$\int\int \cdots \int dx_1\, dx_2 \cdots dx_n$$

For example, variance of a variable x given n independent samples, $\underline{x}$.

$$f(\underline{x}|\alpha) = g(x_1|\alpha)\, g(x_2|\alpha) \cdots g(x_n|\alpha)$$

and $$\alpha = \left\langle \left(x - \langle x \rangle\right)^2 \right\rangle = \langle x^2 \rangle - \langle x \rangle^2$$

$$\langle x \rangle = \int x'\, g(x'|\alpha)\, dx'$$

As an estimator, try:

$$\beta(x) = \overline{(x - \bar{x})^2} = \overline{x^2} - \bar{x}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_i x_i\right)^2$$

$$= \frac{1}{n}\sum_i x_i^2 - \frac{1}{n^2}\left(\sum_i x_i^2 + \sum_{i\neq j} x_i x_j\right)$$

$$= \frac{1}{n^2}\left((n-1)\sum_i x_i^2 - \sum_{i\neq j} x_i x_j\right)$$

So $\langle\beta\rangle = \frac{1}{n^2}\left[(n-1)\sum_i \underbrace{\int x_i^2 g(x_i|\alpha)\,dx_i}_{\langle x^2\rangle} - \sum_{i\neq j} \underbrace{\int x_i g(x_i|\alpha)\,dx_i}_{\langle x\rangle}\underbrace{\int x_j g(x_j|\alpha)\,dx_j}_{\langle x\rangle}\right]$

$$= \frac{1}{n^2}\left[(n-1)\cdot n\langle x^2\rangle - n(n-1)\cdot\langle x\rangle^2\right]$$

$$= \frac{n-1}{n}\underbrace{\left(\langle x^2\rangle - \langle x\rangle^2\right)}_{\alpha}$$

$\therefore \quad \langle\beta\rangle = \frac{n-1}{n}\alpha$. So β is a "biased" estimator.

Easy to construct an "unbiased" estimator $\hat{\beta}$:

$$\hat{\beta} = \frac{n}{n-1}\beta. \quad \text{Then} \quad \langle \hat{\beta} \rangle = \alpha.$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \leftarrow \text{usual "sample variance" formula.}$$

- "degree of freedom" argument
- fishy?

How well do we expect $\hat{\beta}$ to do on average?

← define wellness via "cost" function $(\alpha - \hat{\beta})^2$.

For arbitrary estimator β, (error)

$$\langle (\alpha - \beta)^2 \rangle = \alpha^2 - 2\alpha\langle\beta\rangle + \langle\beta^2\rangle$$

$$\beta(x) = \underbrace{(\alpha - \langle\beta\rangle)^2}_{\text{"bias"}} + \underbrace{\langle\beta^2\rangle - \langle\beta\rangle^2}_{\text{"variance"}}$$

Want this to be small, so want to eliminate both "bias" and "variance"

→ "unbiased minimum variance" estimators

"efficient" in older terminology.

In our example of sample variance, what is the effect on the quality of the estimate by eliminating the bias?

bias is $(\alpha - \langle\beta\rangle)^2 = \alpha^2\left(1 - \left(\frac{n-1}{n}\right)\right)^2$

$= \frac{\alpha^2}{n^2}$, so we lose that from the cost function.

new variance is: $\langle\hat{\beta}^2\rangle - \langle\hat{\beta}\rangle^2 = \left(\frac{n}{n-1}\right)^2\left(\underbrace{\langle\beta^2\rangle - \langle\beta\rangle^2}_{\text{old variance.}}\right)$

For large n, lose small bias contribution,
but variance contribution stays ~ constant,
but gets slightly bigger.

Is this the best we can do?

Why quadratic loss?

Where do estimators come from? Do we have to guess them then measure their cost?

Let's keep the cost function arbitrary: $L(\alpha, \beta)$.

Expected cost: $R_\alpha = \int L(\alpha, \beta)\, f(\underline{x}|\alpha)\, d\underline{x}$

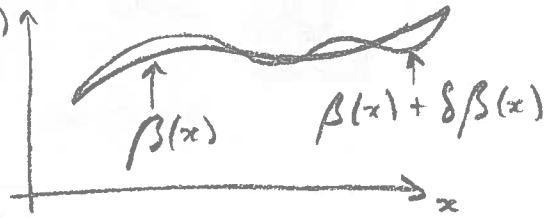
$\beta(\underline{x})$

Given L, what is $\beta(\underline{x})$ that minimises R_α?

Variational calculus!

$$\delta R_\alpha = \int \frac{\delta L(\alpha, \beta(\underline{x}))}{\delta \beta}\, \delta\beta\, f(\underline{x}|\alpha)\, d\underline{x}$$

eg in 1-d:



R_α at a minimum when $\delta R_\alpha = 0 \quad \forall\, \delta\beta$

$$\Rightarrow \quad \frac{\delta L}{\delta \beta} = 0 \quad ?$$

So loss function doesn't depend on β?

- need to rethink!

We are asking: "given α, what is the best estimator?"

answer is: $\beta(\underline{x}) = \alpha$ - indep of $\underline{x}$!

Any solution to the problem as posed would depend on α, but we ... actually, start a new page for this.

Any solution to the problem as posed would depend on the particular value of α one which $f(\underline{x}|\alpha)$ is conditioned,

so: "$\beta_\alpha(\underline{x})$"

But we don't know α - that is why we need an estimator in the first place! The best estimator will protect against cost no matter what α turns out to be.

So instead minimize:

$$R = \int g(\alpha) R_\alpha \, d\alpha$$

where $g(\alpha)$ measures relative importance of R_α for the different possible values of α.

Now: variational calculus again:

$$\delta R = \int_{\underline{x}} \int_{\alpha} g(\alpha) \frac{\delta L(\alpha,\beta)}{\delta \beta} f(\underline{x}|\alpha) \, d\alpha \; \delta\beta(\underline{x}) \, d\underline{x}$$

for this to vanish under arbitrary small changes to $\beta(\underline{x})$,

this must vanish.

$$\int g(\alpha) \frac{\delta L}{\delta \beta}(\alpha, \beta)\, f(\underline{x}|\alpha)\, d\alpha = 0$$

This lets us solve for functional form of β given a choice of cost metric $L(\alpha, \beta)$ that gives the best estimator for α.

Eg: quadratic cost from before: $L(\alpha, \beta) = C(\alpha - \beta)^2$ (C: const)

get: $\int g(\alpha)\,(\alpha - \beta)\, f(\underline{x}|\alpha)\, d\alpha = 0$

$$\Rightarrow \beta(\underline{x}) = \frac{\int \alpha \cdot g(\alpha)\, f(\underline{x}|\alpha)\, d\alpha}{\int g(\alpha)\, f(\underline{x}|\alpha)\, d\alpha}$$

note ordering!

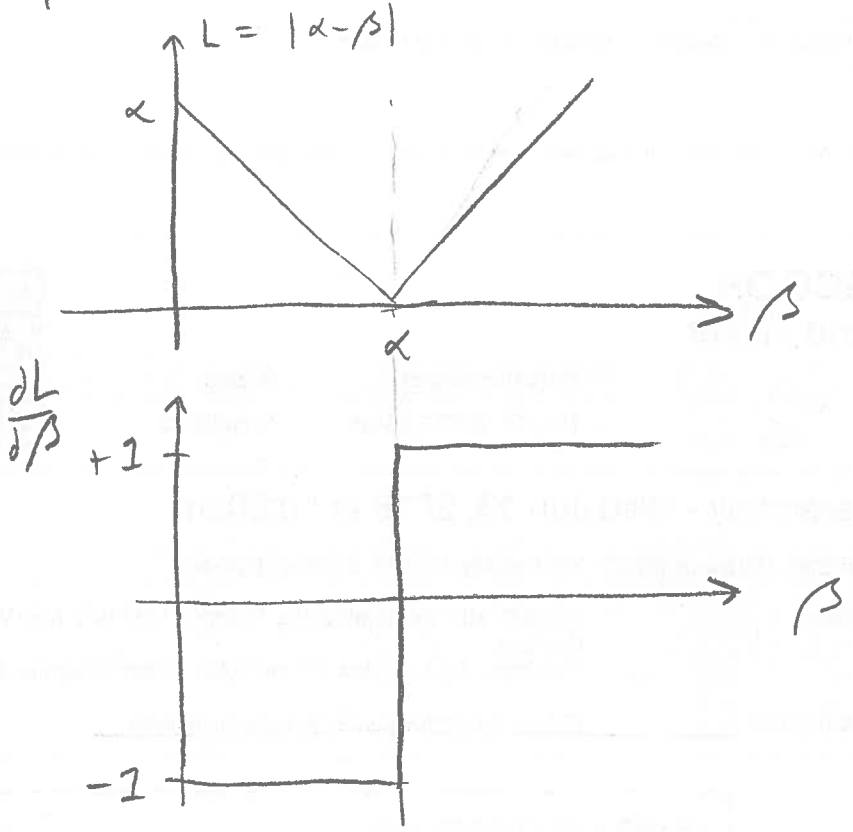
write as $\beta(\underline{x}) = \int \alpha\, f(\alpha|\underline{x})\, d\alpha$

where $f(\alpha|\underline{x}) = \dfrac{g(\alpha)\, f(\underline{x}|\alpha)}{\int g(\alpha)\, f(\underline{x}|\alpha)\, d\alpha}$ is the (normalized) posterior density for α if $g(\alpha)$ is the prior.

ie $\beta(\underline{x}) = \langle \alpha \rangle$, the posterior mean.

over prob of α, not $\underline{x}$, now!

Next example: absolute error $L(\alpha, \beta) = |\alpha - \beta|$.



so in α integral, when $\alpha < \beta$, $L = -1$

$\alpha > \beta$, $L = +1$

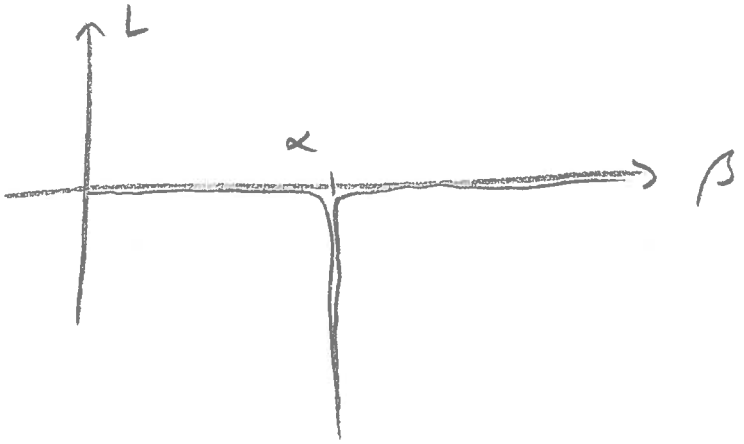
$$\rightarrow \quad \int_{-\infty}^{\beta} g(\alpha)\, f(\underline{x}|\alpha)\, d\alpha = \int_{\beta}^{\infty} g(\alpha)\, f(\underline{x}|\alpha)\, d\alpha$$

$\Rightarrow$ β is <u>median</u> over posterior density $f(\alpha|\underline{x})$.

Final example: L is ~ "inverted delta function"

$$L(\alpha,\beta) = -\delta(\alpha-\beta).$$

So, dramatic drop in cost when β estimates correctly, but if incorrect, don't really care by how much:



Have, in terms of posterior density $f(\alpha|\underline{x})$:

$$\int \frac{\partial L}{\partial \beta} f(\alpha|\underline{x})\, d\alpha = 0$$

$$\therefore -\frac{\partial}{\partial \beta}\int \delta(\alpha-\beta)\, f(\alpha|\underline{x})\, d\alpha = 0$$

$$\therefore \frac{\partial}{\partial \beta} f(\beta|\underline{x}) = 0 \rightarrow \underline{\text{mode}} \text{ of posterior.}$$

If $g(\alpha) \sim$ const where likelihood $f(\underline{x}|\alpha)$ peaks, then we have: Maximum Likelihood Estimate (MLE)

- have insight into what it achieves / when it is appropriate.

ie when

Moral

"Bias" in statistics is a term that carries an implied value judgement, causing us to be biassed against it. By removing "bias" alone, we don't get the best estimator. In that sense, "unbiassed" estimators are biassed away from the best answer. By removing <u>our</u> bias against statistical "bias", we arrive at truly unbiassed estimators - those derived via

$$\int \frac{\delta L}{\delta \beta} \, g(\alpha) \, f(\underline{x}|\alpha) \, d\alpha = 0$$