# Proofs for the Bayes isn't just regularization post

Reginald Reagan aka RAGIN' RAYGUNS

July 16, 2017

## Contents



## 1 Definitions and lemmas used in both unbiasedness proofs

The proof focuses on symmetries to reflections on the real line. A reflection involves shifting to zero, multiplying by negative one, and shifting back:

$$R_a(x) = (x - a) \cdot (-1) + a = 2a - x$$

**Lemma 1.**

$$\mathbb{E}[R_a(X)] = R_a(\mathbb{E}[X])$$

*Proof.*

$$\mathbb{E}[R_a(X)] = \mathbb{E}[2a - X] = 2a - \mathbb{E}[X] = R_a(\mathbb{E}[X])$$

□

In the above you can think of $x$ as a number. However, below $x$ will be a vector, specifically a vector of observations in the sample. I will use the symbol $R_a$ for both the above operation on a number, as well as the same operation applied elementwise to a vector.

**Lemma 2.** *If a sample $X$ is drawn from a Cauchy distribution centered on $\mu$, then for any function $f$,*

$$\mathbb{E}_\mu[f(R_\mu(X))] = \mathbb{E}_\mu[f(X)]$$

*Proof.*

$$\int_{-\infty}^{\infty} f(R_\mu(x)) \prod_i \pi(1+(x_i-\mu)^2)^{-1} dx$$

Changing variables with $y_i = R_\mu(x_i)$,

$$\int_{\infty}^{-\infty} f(y) \prod_i \pi(1+(2\mu-y_i-\mu)^2)^{-1}(-1) dy$$

$$\int_{-\infty}^{\infty} f(y) \prod_i \pi(1+(y_i-\mu)^2)^{-1} dy$$

□

## 2 Posterior mean is unbiased

**Definition 1.** *The posterior mean $T(x)$ is this function of the sample $x$:*

$$T(x) = \int_{-\infty}^{\infty} \mu' \frac{\prod_i \pi(1+(x_i-\mu')^2)^{-1}}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\mu'')^2)^{-1} d\mu''} 1 d\mu'$$

I'm writing it this way so you can see that the part after $\mu'$ is the posterior distribution, written using Bayes' theorem. The top is the conditional probability of the data given $\mu'$, the bottom is the marginal probability of the data, and the 1 is our improper uniform prior. But here's a better way to write it:

$$T(x) = \frac{\int_{-\infty}^{\infty} \mu' \prod_i \pi(1+(x_i-\mu')^2)^{-1} d\mu'}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\mu')^2)^{-1} d\mu'}$$

This is the form I'm going to use below.

**Lemma 3.** *For any $a$,*

$$R_a(T(x)) = T(R_a(x))$$

*Proof.*

$$T(x+b) = \frac{\int_{-\infty}^{\infty} \mu' \prod_i \pi(1+(x_i+b-\mu')^2)^{-1} d\mu'}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i+b-\mu')^2)^{-1} d\mu'}$$

Changing variables with $\nu = \mu' - b$,

$$T(x+b) = \frac{\int_{-\infty}^{\infty} (\nu+b) \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu}$$

$$= \frac{\int_{-\infty}^{\infty} \nu \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu} + b \frac{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu}$$

$$= \frac{\int_{-\infty}^{\infty} \nu \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu}{\int_{-\infty}^{\infty} \prod_i \pi(1+(x_i-\nu)^2)^{-1} d\nu} + b$$

$$T(x+b) = T(x) + b$$

$$T(-x) = \frac{\int_{-\infty}^{\infty} \mu' \prod_i \pi(1 + (-x_i - \mu')^2)^{-1} d\mu'}{\int_{-\infty}^{\infty} \prod_i \pi(1 + (-x_i - \mu')^2)^{-1} d\mu'}$$

Changing variables with $\nu = -\mu'$:

$$\frac{\int_{\infty}^{-\infty} -\nu \prod_i \pi(1 + (-x_i + \nu)^2)^{-1} (-1) d\nu}{\int_{\infty}^{-\infty} \prod_i \pi(1 + (-x_i + \nu)^2)^{-1} (-1) d\nu}$$

$$= -\frac{\int_{-\infty}^{\infty} \nu \prod_i \pi(1 + (x_i - \nu)^2)^{-1} d\nu}{\int_{-\infty}^{\infty} \prod_i \pi(1 + (x_i - \nu)^2)^{-1} d\nu}$$

$$T(-x) = -T(x)$$

$$T(R_a(x)) = T(2a - x) = 2a + T(-x) = 2a - T(x) = R_a(T(x))$$

□

**Theorem 1** (Posterior mean is unbiased)**.**

$$\mathbb{E}_\mu[T(X)] = \mu$$

*Proof.*

$$\begin{array}{ll} R_\mu(\mathbb{E}_\mu[T(X)]) & \\ = \mathbb{E}_\mu[R_\mu(T(X))] & \text{Lemma 1} \\ = \mathbb{E}_\mu[T(R_\mu(X))] & \text{Lemma 3} \\ = \mathbb{E}_\mu[T(X)] & \text{Lemma 2} \end{array}$$

Summarizing the above,

$$R_\mu(\mathbb{E}_\mu[T(X)]) = \mathbb{E}_\mu[T(X)]$$

$$2\mu - \mathbb{E}_\mu[T(X)] = \mathbb{E}_\mu[T(X)]$$

$$2\mu = 2\mathbb{E}_\mu[T(X)]$$

$$\mu = \mathbb{E}_\mu[T(X)]$$

□

## 3 Maximum likelihood estimate is unbiased

**Definition 2.** *The maximum likelihood estimate is this function of the sample $x$:*

$$U(x) = \operatorname*{argmax}_{\mu'} \prod_i \pi(1 + (x_i - \mu')^2)^{-1}$$

**Lemma 4.** *For any $a$,*

$$R_a(U(x)) = U(R_a(x))$$

*Proof.* Let $\lambda(\mu')$ be the likelihood function for the sample $x$:

$$\lambda(\mu') = \prod_i \pi(1 + (x_i - \mu')^2)^{-1}$$

Then,

$$U(x) = \operatorname*{argmax}_{\mu'} \lambda(\mu')$$

Let $\lambda'(\mu')$ be the likelihood function for the sample $R_a(x)$:

$$\lambda'(\mu') = \prod_i \pi(1 + (R_a(x_i) - \mu')^2)^{-1}$$

Then,

$$U(R_a(x)) = \operatorname*{argmax}_{\mu'} \lambda'(\mu')$$

Connecting the values of these two likelihood functions:

$$\lambda'(R_a(\mu')) = \prod_i \pi(1 + (R_a(x_i) - R_a(\mu'))^2)^{-1}$$

$$= \prod_i \pi(1 + (2a - x_i - (2a - \mu'))^2)^{-1}$$

$$= \prod_i \pi(1 + (x_i - \mu')^2)^{-1}$$

$$= \lambda(\mu')$$

Let $y^* = \max_{\mu'} \lambda(\mu')$. Every value taken by $\lambda'(\mu')$ is taken by $\lambda(\mu')$ somewhere, so we also have $y^* = \max_{\mu'} \lambda'(\mu')$. We know that

$$\lambda(U(x)) = y^*$$

Therefore,

$$\lambda'(R_a(U(x))) = \lambda(U(x)) = y^*$$

$$R_a(U(x)) = \operatorname*{argmax}_{\mu'} \lambda(\mu') = U(R_a(x))$$

□

**Theorem 2** (Maximum likelihood estimator is unbiased)**.**

$$\mathbb{E}_\mu[T(X)] = \mu$$

*Proof.* Same proof as Theorem 1, but use lemma 4 instead of lemma 3. □

# 4 Location equivariance of the posterior mean and maximum likelihood estimates

As before, $T(x)$ is the posterior mean calculated from the sample $x$, and $U(x)$ is the maximum likelihood estimate. See sections 2 and 3 respectively for the full definitions. The proof that $T(x+b) = T(x)+b$ is actually back in section 2, as part of the proof of lemma 3. The proof that $U(x+b) = U(x)+b$ is a slight variation on the proof of lemma 4, but I'll spell it out in its entirety below.

**Theorem 3.** *For any $b$,*

$$U(x+b) = U(x) + b$$

*Proof.* Let $\lambda(\mu')$, as before, be the likelihood function for the sample $x$:

$$\lambda(\mu') = \prod_i \pi(1 + (x_i - \mu')^2)^{-1}$$

Then,

$$U(x) = \operatorname*{argmax}_{\mu'} \lambda(\mu')$$

Let $\lambda'(\mu')$, this time, be the likelihood function for the sample $x+b$:

$$\lambda'(\mu') = \prod_i \pi(1 + (x + b - \mu')^2)^{-1}$$

Then,

$$U(x+b) = \operatorname*{argmax}_{\mu'} \lambda'(\mu')$$

Connecting the values of these two likelihood functions:

$$\lambda'(\mu' + b) = \prod_i \pi(1 + (x + b - (\mu' + b))^2)^{-1}$$

$$= \prod_i \pi(1 + (x - \mu')^2)^{-1}$$

$$= \lambda(\mu')$$

Let $y^* = \max_{\mu'} \lambda(\mu')$. Every value taken by $\lambda'(\mu')$ is taken by $\lambda(\mu')$ somewhere, so we also have $y^* = \max_{\mu'} \lambda'(\mu')$. We know that

$$\lambda(U(x)) = y^*$$

Therefore,

$$\lambda'(U(x) + b) = \lambda(U(x)) = y^*$$

$$U(x) + b = \operatorname*{argmax}_{\mu'} \lambda(\mu') = U(x+b)$$

□