KMart as a Bayesian audit

Damjan Vukcevic

25 Jun 2019 (updated 9 Dec 2019)

This note shows a proof that KMart (the martingale approach described in this repository) is equivalent to a Bayesian audit with a risk-maximising uniform prior for the reported winner's true vote tally. It also introduces a more general version of the test statistic that corresponds to an arbitrary risk-maximising prior. Both results are shown for the case of sampling with replacement, for a simple 2-candidate election.

1 KMart is equivalent to a Bayesian audit

Suppose we are auditing a simple 2-candidate election, using sampling with replacement. We observe iid $X_1, X_2, \ldots \in \{0, 1\}$, where $X_j = 1$ is a vote for the reported winner and $X_j = 0$ is a vote for the reported loser. Let $\mathbb{E}X_j = t$, the true tally of the reported winner. In other words, the X_j are a sequence of Bernoulli trials with success probability t.

The null hypothesis for the audit is that the reported winner actually lost, i.e. that $t \leqslant \frac{1}{2}$. To carry out a test, we usually set this to the 'hardest' case[1], which is $H_0 \colon t = t_0 = \frac{1}{2}$. The alternative hypothesis is that the winning candidate was reported correctly, i.e. $H_1 \colon t > \frac{1}{2}$.

In practice we will always have a finite number of total votes, and thus a realistic model would have the support of t be a discrete set (i.e. values of the form k/N where N is the total number of votes). However, for mathematical convenience here we will allow the support of t to be the unit interval, which is continuous.

1.1 KMart audits

KMart is a risk-limiting election auditing method based on martingale theory. For the context described above, it uses the following test statistic:

$$A_n = \int_0^1 \prod_{j=1}^{n} \left(\gamma \left[\frac{X_j}{t_0} - 1 \right] + 1 \right) d\gamma.$$

Since we are working with $t_0 = \frac{1}{2}$, we can rewrite this expression,

$$A_n = 2^n \int_0^1 \prod_{j=1}^{n} \left(\gamma \left[X_j - \frac{1}{2} \right] + \frac{1}{2} \right) d\gamma.$$

For a specified risk limit, α, the audit proceeds until $A_n > 1/\alpha$, at which point the election is certified (H_0 is rejected), or is otherwise terminated in favour of doing a full recount.

[1] 'Hardest' means that it is the case that leads to the largest false positive rate (miscertification probability), i.e. the *risk*.

1.2 Bayesian audits

A Bayesian audit is based on standard Bayesian inference. The verdict of the audit is based on the posterior probability that the reported winner actually won (or lost, in which case this is called the *upset probability*). Typically, a threshold will be placed on this probability for deciding whether to certify the election or carry on sampling.[2]

Bayesian audits can be represented in terms of the posterior odds, which gives a similar formulation to other risk-limiting audits (Vora, 2019). For the context described above, they would use the following test statistic:

$$B_n = \frac{\Pr(H_1 \mid X_1, \ldots, X_n)}{\Pr(H_0 \mid X_1, \ldots, X_n)} = \frac{\Pr(X_1, \ldots, X_n \mid H_1)}{\Pr(X_1, \ldots, X_n \mid H_0)} \times \frac{\Pr(H_1)}{\Pr(H_0)}.$$

We will limit our discussion to risk-maximising prior distributions[3]. These place a probability mass of $\frac{1}{2}$ on the value of $t = \frac{1}{2}$, and the remaining probability is over the set $t \in (\frac{1}{2}, 1]$. That means that $\Pr(H_1) = \Pr(H_0) = \frac{1}{2}$, meaning that the prior odds drop out of the above equation. The remaining term is the Bayes factor (BF). Let's write this out more explicitly.

Let $Y_n = \sum_{j=1}^{n} X_n$. The denominator of the BF is simple: the likelihood of the sample at the (point) null value,

$$\Pr(X_1, \ldots, X_n \mid H_0) = \Pr\left(X_1, \ldots, X_n \mid t = \frac{1}{2}\right) = \left(\frac{1}{2}\right)^{Y_n} \left(\frac{1}{2}\right)^{n-Y_n} = \frac{1}{2^n}.$$

The numerator requires integrating over the prior under H_1. Letting this be $f(t)$, where $t \in (\frac{1}{2}, 1]$, allows us to write the numerator as,

$$\Pr(X_1, \ldots, X_n \mid H_1) = \int_{\frac{1}{2}}^{1} t^{Y_n} (1-t)^{n-Y_n} f(t)\, dt.$$

Putting these together gives,

$$B_n = 2^n \int_{\frac{1}{2}}^{1} t^{Y_n} (1-t)^{n-Y_n} f(t)\, dt.$$

Similar to KMart, a Bayesian audit proceeds until $B_n < 1/\alpha$.

1.3 Equivalence

Both A_n and B_n are expressed as integrals but with the X_j in different 'places' in the integrand. The key to showing they are equivalent is to notice that the X_j are binary variables, which allows us to set up an identity that relates the two ways of writing the integral. Specifically, we have the following identity,

$$\gamma\left(X_j - \frac{1}{2}\right) + \frac{1}{2} = \left(\frac{1+\gamma}{2}\right)^{X_j} \left(\frac{1-\gamma}{2}\right)^{1-X_j}.$$

This allows us to rewrite A_n,

$$A_n = 2^n \int_0^1 \left(\frac{1+\gamma}{2}\right)^{Y_n} \left(\frac{1-\gamma}{2}\right)^{n-Y_n} d\gamma = \int_0^1 (1+\gamma)^{Y_n} (1-\gamma)^{n-Y_n}\, d\gamma.$$

[2] A Bayesian audit requires specifying a prior on the winner's vote tally. For any given prior, it is possible to specify a threshold on the posterior probability such that the audit limits the risk. Alternatively, it is possible to select priors such that the threshold itself is the risk limit. Further details are in Vora (2019) and a note by Vukcevic in a separate Git repo.

[3] See Vora (2019) for an example with a discrete support.

Next, let $\gamma = 2t - 1$ and change the variable of integration,

$$A_n = \int_{\frac{1}{2}}^{1} (2t)^{Y_n} (2 - 2t)^{n - Y_n} 2 \, dt = 2^n \int_{\frac{1}{2}}^{1} t^{Y_n} \, (1 - t)^{n - Y_n} \, 2 \, dt.$$

Finally, note that this is identical to B_n if we set the prior to be uniform over H_1, i.e. $f(t) = 2$.

In other words, a KMart audit is equivalent to a Bayesian audit that uses a risk-maximising uniform prior.

2 Extending KMart to arbitrary priors

From the above result, we can see that γ plays a similar role to t. The somewhat arbitrary integral over γ used to define A_n can be generalised by specifying a weighting function $g(\gamma)$,

$$A_n = \int_0^1 \prod_{j=1}^{n} \left(\gamma \left[\frac{X_j}{t_0} - 1 \right] + 1 \right) g(\gamma) \, d\gamma.$$

Applying the same transformations as above gives,

$$A_n = 2^n \int_{\frac{1}{2}}^{1} t^{Y_n} \, (1 - t)^{n - Y_n} \, 2 \times g(2t - 1) \, dt.$$

In other words, this generalised version of KMart is equivalent to a Bayesian audit with the following risk-maximising prior:

$$f(t) = 2 \times g(2t - 1).$$

The original KMart is the special case where $g(\cdot) = 1$.

3 Discussion

This equivalence result sheds light on the nature of the KMart method. It is equivalent to a (risk-limiting) Bayesian audit with a specific prior. The extension described here allows use of an arbitrary prior.

Thus, we can think of KMart as providing an alternative expression for risk-limiting Bayesian audits. Perhaps this expression has some computational advantages?

In terms of relative performance, the question can therefore be reduced to asking about the relative performance of different prior distributions. As a starting point, the 'answer' will be somewhat boring: the methods that perform best will be simply the ones for which the prior most closely matches the truth. (However, we need to be clear exactly how we measure performance. . .)

For this reason, it is of interest to know how the different priors behave across a range of scenarios, i.e. how sensitive/robust they are. We probably want to use priors that give relatively robust results.

I suspect that the uniform prior is going to be quite robust and a sensible choice.

As a possible improvement in performance (but at the expense of robustness?), we could try using priors that incorporate some knowledge of the election. For example, suppose the reported tally for the reported winner is $t^\star$. Under H_1, we could use a uniform prior over $(\frac{1}{2}, t^\star]$. Or, more ambitiously, we could specify a narrow interval around $t^\star$ and place all of our prior mass on that, which we can think of as a 'fuzzy' version of BRAVO.

4 Efficient computation by exploiting the equivalence

We can use the above equivalence to develop fast ways to compute the KMart statistic, by relating it to standard Bayesian calculations using conjugate priors.

First, we show that if we take a conjugate prior distribution, truncate it, and add some point masses, the resulting distribution is still conjugate. Then we use this result to write a formula for the posterior distribution for the same case as above (simple 2-candidate election, sampling with replacement).

4.1 Truncation and point masses preserve conjugacy

(The proofs shown here are not too hard to derive and may well be described elsewhere.)

Suppose we have a single parameter, θ, some data, D, a likelihood function, $L(\theta \mid D)$, and a conjugate prior distribution, $f(\theta)$. That means we have,

$$f(\theta \mid D) \propto L(\theta \mid D) f(\theta).$$

Let the normalising constant be,

$$k = \int L(\theta \mid D) f(\theta) d\theta.$$

This allows us to express the posterior as,

$$f(\theta \mid D) = \frac{1}{k} L(\theta \mid D) f(\theta),$$

The sections that follow each start with these definitions and transform the prior in various ways.

4.1.1 Truncation

Truncate the prior to a subset S (i.e. we only allow $\theta \in S$). Write this truncated prior as,

$$f^*(\theta) = f(\theta) \frac{I_S(\theta)}{z_S},$$

where $I_S(\theta)$ is the indicator function that takes value 1 when $\theta \in S$, and $z_S = \int f(\theta) I_S(\theta) d\theta = \int_S f(\theta) d\theta$ is the normalising constant due to truncation.

If we use this prior, what posterior do we get? It will be,

$$f^*(\theta \mid D) = \frac{1}{k^*} L(\theta \mid D) f^*(\theta),$$

where,

$$k^* = \int L(\theta \mid D) f^*(\theta) d\theta.$$

Expanding this out gives,

$$f^*(\theta \mid D) = \frac{1}{k^* z_S} L(\theta \mid D) f(\theta) I_S(\theta) = \frac{k}{k^* z_S} f(\theta \mid D) I_S(\theta).$$

This is the original posterior truncated to S. Thus, the truncation results in staying within the same family of (truncated) probability distributions, which means this family is conjugate.

4.1.2 Adding a point mass

Define a 'spiked' prior where we add a point mass at θ_0,

$$f^*(\theta) = a\,\delta_{\theta_0}(\theta) + bf(\theta),$$

where $a + b = 1$. In other words, a mixture distribution with mixture weights a and b. The normalising constant is,

$$k^* = \int L(\theta \mid D) f^*(\theta) d\theta = aL(\theta_0 \mid D) + bk.$$

We can write the posterior as,

$$f^*(\theta \mid D) = \frac{1}{k^*} L(\theta \mid D) f^*(\theta) = \frac{a\,L(\theta_0 \mid D)}{k^*}\delta_{\theta_0}(\theta) + \frac{bk}{k^*} f(\theta \mid D).$$

This is a 'spiked' version of the original posterior. You can see this more clearly by defining,

$$a^* = \frac{a\,L(\theta_0 \mid D)}{k^*}, \quad b^* = \frac{bk}{k^*},$$

where $a^* + b^* = 1$. Thus, 'spiking' a distribution results in a conjugate family. Note that the mixture weights get updated as we go from the prior to the posterior.

4.1.3 Truncating and adding point masses

We can combine both of the previous operations and we will still retain conjugacy. In fact, due to the generality of the proof, we can apply each one an arbitrary number of times, e.g. to add many point masses.

4.2 Application to KMart

When sampling with replacement, the conjugate prior for t (the true tally of the reported winner) is a beta distribution.

We showed earlier that KMart was equivalent to using a risk-maximising prior. Starting with any beta distribution, we can form the corresponding risk-maximising prior by truncating to $t \in (\frac{1}{2}, 1]$ and adding a probability mass of $\frac{1}{2}$ at $t = \frac{1}{2}$. Based on the argument presented above, this prior is conjugate. Moreover, we can express the posterior in closed form.

Let the original prior be $t \sim \text{Beta}(\alpha, \beta)$. Note that this α is just a hyperparameter and not a specified risk limit. The risk-maximising prior retains the functional form of this prior for $t > \frac{1}{2}$ and also has a mass of $\frac{1}{2}$ at $t = \frac{1}{2}$.

After we observe a sample of size n from the audit, we have a posterior with an updated probability mass at $t = \frac{1}{2}$. This mass will be the upset probability. We can derive an expression for it using equations similar to above (it will correspond to a^* using the notation from above).

Let $f(t)$ be the pdf of the original beta prior, $F(t)$ be its cdf, $S = (\frac{1}{2}, 1]$ the truncation region, $F'(t)$ the cdf of the beta-distributed portion of the posterior (i.e. the posterior distribution if we use the original beta prior), and $B(\cdot, \cdot)$ be the beta function. We have,

$$k^* = \frac{1}{2}\left(\frac{1}{2}\right)^n + \frac{1}{2}\frac{k'}{z_S},$$

where

$$z_S = \int_{\frac{1}{2}}^{1} f(t)dt = 1 - F\left(\frac{1}{2}\right)$$

and

$$k' = \int_{\frac{1}{2}}^{1} L(t \mid D) f(t)dt = \frac{B(Y_n + \alpha, n - Y_n + \beta)}{B(\alpha, \beta)} \left(1 - F'\left(\frac{1}{2}\right)\right).$$

Putting these together gives,

$$k^* = \frac{1}{2^{n+1}} + \frac{1}{2} \times \frac{B(Y_n + \alpha, n - Y_n + \beta)}{B(\alpha, \beta)} \times \frac{1 - F'\left(\frac{1}{2}\right)}{1 - F\left(\frac{1}{2}\right)}.$$

The upset probability is,

$$a^* = \frac{\frac{1}{2^{n+1}}}{k^*}.$$

These quantities will be straightforward to calculate as long we have efficient ways to calculate:

1. The beta function
2. The cdf of a beta distribution

Both have fast implementations in R.