A note on using Concentration Inequalities for sampling without replacement to bound future rewards

Mark Alexander Burgess

Australian National University, Canberra, ACT, 2600, Australia

Abstract

We show how some concentration inequalities for sampling without replacement can be used for bounding future samples. This process can be extended to bound the sum of future samples from multiple populations, and we analyse an illustrative sample allocation problem.

Keywords: concentration bounds, sampling without replacement

Many famous concentration inequalities are Chernoff bounds which extend from the creation of upper bounds for the moment generating function of the random variable in question. Usually concentration inequalities are developed in the context of sampling with replacement, but there are some which utilise martingale arguments to give additional refinement in the context of sampling without replacement. The first and most notable martingale argument along these lines was given by Serfling [1974], but recent refinements of this argument have been given by Riggs et al. [2013] and Bardenet and Maillard [2015].

What may not be obvious, is how these concentration inequalities which are developed to bound the sample mean in the context of sampling without replacement, can be used for bounding future samples in the context of sampling with replacement, as we will show. The simple idea is to consider the past and future samples in the context of sampling without replacement as if it were a finite set, and then using concentration inequalities for sampling without

Email address: `mark.burgess@anu.edu.au` (Mark Alexander Burgess)

replacement to characterise the future samples in terms of the past samples without direct reference to the mean of the finite set.

We will be using an example upper bound for the moment generating function developed by Bardenet and Maillard [2015][Proposition 2.3]

Theorem 1. *for i.i.d random variables $X_{1\leq,i\leq N}$, with average value $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$, which are bounded $a \leq X_i \leq b$, with $Z_k = \frac{1}{k}\sum_{t=1}^{k}(X_t - \mu)$ as the average of the first k of them above the mean, then for $\lambda > 0$:*

$$\log \mathbb{E} \exp(\lambda n Z_n) \leq \frac{(b-a)^2}{8}\lambda^2(n+1)\left(1 - \frac{n}{N}\right)$$

Bardenet and Maillard developed this bound in the context of sampling specifically without replacement, but the proof relies only on the i.i.d of the samples X_i and the fact that their average is μ (see Appendix A). We can convert this into a bound for future rewards as follows:

Theorem 2. *For n independant samples of a random variable X, X_i (that is bounded $a \leq X \leq b$) if A_n is the average value of thoes samples, for any $y \geq 0$, the sum of m future samples S is probability bounded:*

$$\mathbb{P}(S \leq mA_n - y) \leq \exp\left(\frac{-2y^2n^2}{(b-a)^2(m+n)(n+1)m}\right)$$

Proof. In the context of Theorem 1, if $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the average of the first n values, and $S = \sum_{i=n+1}^{N} X_i$ is the sum of the remaining $N - n$ terms, then $Z_n = (1 - n/N)A_n - S/N$, hence:

$$\log \mathbb{E} \exp(\lambda n((1-n/N)A_n - S/N)) \leq (b-a)^2\lambda^2(n+1)\left(1 - n/N\right)/8$$

scaling $\lambda \to \lambda N/n$ and substituting $N = m + n$ gives:

$$\log \mathbb{E} \exp(\lambda(mA_n - S)) \leq (b-a)^2\lambda^2(m+n)(n+1)mn^{-2}/8 \tag{1}$$

We use this bound to create a Chernoff bound by Markov's inequality:

$$\mathbb{P}(mA_n - S \geq y) = \mathbb{P}(\exp(\lambda(mA_n - S)) \geq \exp(\lambda y))$$

$$\leq E[\exp(\lambda(mA_n - S))]\exp(-\lambda y) = \exp((b-a)^2\lambda^2(m+n)(n+1)mn^{-2}/8 - \lambda y)$$

Minimising λ, at $\lambda = 4n^2y((b-a)^2(m+n)(n+1)m)^{-1}$ giving the result. □

We now consider an extension and example problem for sampling from many distributions:

Theorem 3. *For n_j independent samples of random variables X_j (each is bounded $a_j \leq X_j \leq b_j$) if A_{j,n_j} is the average value of the n_j samples of random variable X_j, the sum S of all the m_j future samples from each of the X_j random variables is probability bounded. For any $y \leq \sum_j m_j A_{j,n_j}$:*

$$\mathbb{P}\left(S \leq y\right) \leq \exp\left(\frac{-2(\sum_j m_j A_{j,n_j} - y)^2}{\sum_j \frac{(b_j-a_j)^2}{n_j^2}(m_j+n_j)(n_j+1)m_j}\right)$$

Proof. considering equation 1 for any random variables X_j, then the sum of its future samples S_j is:

$\log \mathbb{E} \exp(\lambda(m_j A_n - S_j)) \leq (b_j - a_j)^2 \lambda^2 (m_j + n_j)(n_j + 1) m_j n_j^{-2}/8$

Since the sampling of each of the random variables is independent then:

$\log \mathbb{E} \exp(\lambda \sum_j m_j A_n - \lambda S) \leq \sum_j (b_j - a_j)^2 \lambda^2 (m_j + n_j)(n_j + 1) m_j n_j^{-2}/8$

Using Markov's inequality in a similar way to the proof of Theorem 1 gives:

$\mathbb{P}(\sum_j m_j A_{j,n_j} - S \geq z) \leq \exp(\lambda^2 \sum_j (b_j - a_j)^2 (m_j + n_j)(n_j + 1) m_j n_j^{-2}/8 - \lambda z)$

Minimising λ occurs at $\lambda = z(\sum_j (b_j - a_j)^2 (m_j + n_j)(n_j + 1) m_j n_j^{-2}/4)^{-1}$ which then gives the required result after substituting $y = \sum_j m_j A_{j,n_j} - z$. □

Example 1. A small impoverished community is ordering test kits for a disease, the community has a budget for 55 new test kits. There are multiple suppliers and each have a different track record:

- Supplier A has supplied 30 test kits in the past, 10 of which were faulty.
- Supplier B has supplied 18 test kits in the past, 5 of which were faulty.
- Supplier C has supplied 50 test kits in the past, 17 of which were faulty.

How many test kits should be bought from each of the suppliers to ensure the likelihood of getting less than 32 non-faulty kits is less than 40%?

We can solve this problem by considering Bernoulli random variables, and running the possible integers m_j to minimise the bound of Theorem 3.
This occurs when $m_A = 14, m_B = 22, m_C = 19$ and $\mathbb{P}\left(S \leq 31\right) \leq 0.39911$
This might not be the ideal way of solving this particular problem, but it perhaps an unorthodox use for concentration inequalities designed for sampling without replacement.

Appendix A. A summary proof of Bardenet and Maillard's bound

We present Hoeffding's Lemma before giving a short proof of Theorem 1.

Lemma 1 (Hoeffding's Lemma). *For a random variable Y that has a mean of zero and is of finite support on the interval $a \leq Y \leq b$, with width $D = b - a$, and for any $s > 0$:* $\mathbb{E}\left[\exp(sY)\right] \leq \exp\left(\frac{1}{8}D^2 s^2\right)$.

Lemma 2 (Martingale Step). *For i.i.d random variables Y_i, and $Z_k = \frac{1}{k}\sum_k^{i=1} Y_i$ is the average of the first k of them, then:* $\mathbb{E}[Y_{k+1}|Z_{k+1}\dots Z_N] = Z_{k+1}$

Proof. Consider that $Y_1, \dots, Y_{k+1}$ are all i.i.d variables equally constrained by specification of the $Z_{k+1}\dots Z_N$ thus for all $j < k+1$:
$\mathbb{E}[Y_{k+1}|Z_{k+1}\dots Z_N] = \mathbb{E}[Y_j|Z_{k+1}\dots Z_N]$
thus $\mathbb{E}[Y_{k+1}|Z_{k+1}\dots Z_N] = \frac{1}{k+1}\sum_{j=1}^{k+1}\mathbb{E}[Y_j|Z_{k+1}\dots Z_N]$
$= \mathbb{E}[\frac{1}{k+1}\sum_{i=1}^{k+1} Y_i|Z_{k+1}\dots Z_N] = \mathbb{E}[Z_{k+1}|Z_{k+1}\dots Z_N] = Z_{k+1}$ □

Proof of Theorem 1. Consider letting $Y_i = X_i - \mu$ then
$Z_k = \frac{1}{k}\sum_{i=1}^{k} Y_i = Z_{k+1} + \frac{1}{k}(Z_{k+1} - Y_{k+1})$
$= (Z_k - Z_{k+1}) + (Z_{k+1} - Z_{k+2}) + \dots + (Z_{N-1} - Z_N)$
$= \frac{1}{k}(Z_{k+1} - Y_{k+1}) + \frac{1}{k+1}(Z_{k+2} - Y_{k+2}) + \dots + \frac{1}{N-1}(Z_N - Y_N)$
Thus: $\exp(sZ_n) = \prod_{k=n}^{N-1}\exp\left(\frac{s}{k}(Z_{k+1} - Y_{k+1})\right)$
By repeated application of the Law of total expectation[1] we get:
$\mathbb{E}[\exp(sZ_n)] = \mathbb{E}\left[\prod_{k=n}^{N-1}\mathbb{E}\left[\exp\left(\frac{s}{k}(Z_{k+1} - Y_{k+1})\right)|Z_{k+1}\dots Z_N\right]\right]$
Because $\mathbb{E}[Y_{k+1}|Z_{k+1}\dots Z_N] = Z_{k+1}$ by Lemma 2, then $Z_{k+1} - Y_{k+1}$ is a random variable with a mean of zero bounded within a width $b - a$ and hence amenable for Hoeffding's Lemma 1 giving: $\mathbb{E}[\exp(sZ_n)] \leq \exp\left((b-a)^2 s^2 \sum_{k=n}^{N-1}\frac{1}{8k^2}\right)$
Using the approximation effectively utilised by Bardenet and Maillard [2015]:
$\sum_{k=n}^{N-1}\frac{1}{k^2} \leq (n+1)(1-n/N)n^{-2}$ and letting $\lambda = sn$ gives the required result. □

Acknowledgements

A great thankyou to my supervisors: Archie Chapman, Paul Scott and Sylive Thiébaux.

[1] that $\mathbb{E}[A] = \mathbb{E}[\mathbb{E}[A|B]]$ hence for function f that $\mathbb{E}[Af(B)] = \mathbb{E}[\mathbb{E}[A|B]f(B)]$, and so on

Declaration of interests

The author declares no conflict of interests in the creation of this research.

References

R. Bardenet and O.-A. Maillard. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 08 2015.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, Mar 1963.

K. E. Riggs, D. M. Young, and J. J. Becnel. Probability inequalities for the sum of random variables when sampling without replacement. *International Journal of Statistics and Probability*, 2:75, 2013.

R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, 2(1):39–48, 01 1974.