Efficiently computing sample expectiles

June 21, 2020

1   Expectile statistics

It is well known that the expected value μ_X of a random variable X with finite second moment minimises the expected squared distance from the random variable. That is,

$$\mu_X = \arg\min_{\mu} \mathbb{E}\left[(X - \mu)^2\right]. \tag{1}$$

Expectiles are a class of summary statistics generalising the expected value [1, 2]. Given an asymmetry parameter $\tau \in (0, 1)$, the τ-expectile, $\epsilon_X(\tau)$ of X is the minimiser of an *asymmetric* version of this expected squared distance, weighting squared positive distances by τ and squared negative distances by $1 - \tau$, as in (2)[1].

$$\epsilon_X(\tau) = \arg\min_{\epsilon} \mathbb{E}\left[[\![X > \epsilon]\!]_{1-\tau}^{\tau}(X - \epsilon)^2\right] \tag{2}$$

The expected value μ_X is recovered as the 0.5-expectile. Expectiles with $\tau > 0.5$ are more sensitive to 'higher-than-expected' outcomes than they are to 'lower-than-expected' outcomes. In this sense, they can be viewed as 'optimistic' expectation values. Likewise, expectiles with $\tau < 0.5$ can be viewed as 'pessimistic' expectations. We can thus understand τ as capturing the degree of 'positive outlook' of an expectile, on a scale from 0 to 1.

Expectiles have been relatively neglected as summary statistics since their introduction [3]. Perhaps this is because expectiles lacked a simple calculation procedure for samples, and an immediate interpretation, like those of the expected value or those of quantiles. Viewing expectiles as 'expectations at varying degrees of positive outlook' may provide some interpretability. We outline an efficient method for computing sample expectiles in the next section.

Quantiles and expectiles are closely related. Quantiles minimise an asymmetric *absolute* distance and thus expectiles generalise the expected value in analogy to how quantiles generalise the *median* value of a random variable. Expectiles possess many properties similar to quantiles. Indeed, [4] showed that the expectiles of a random variable X *are* the quantiles of a suitably transformed X.

---

[1] Here, $[\![P]\!]_b^a$ denotes a *generalised Iverson bracket*, evaluating to a if P is a true proposition, or to b otherwise.

Furthermore, expectiles may be preferred over quantiles in several contexts. In regression (i.e. in estimating the distribution of a dependent variable conditional on some independent variable), asymmetric least squares regression (fitting expectile curves) may be preferable to asymmetric absolute value regression (fitting quantile curves) [5]. Based on a correspondence between quantiles and expectiles, [6] suggests an efficient method for estimating quantiles from estimated expectiles, rather than directly from data. This same idea underlies a technique using a distribution's expectiles to compute the *value at risk* and the *conditional value at risk* (a.k.a. *expected shortfall*) [7], establishing expectiles as a sensible learning target for risk-sensitive reinforcement learning, for example [8].

2 Algorithm for sample expectiles

To our knowledge, an efficient method of computing the expectiles of a sample is not available within standard software libraries. Here we derive a method for computing the expectiles of a sample without resorting to iterative optimisation, based on the observation that the sample expectile optimisation target has a piece-wise-linear continuous gradient for which the root may be easily computed.

Given a sample $\vec{x} = x_1, \ldots, x_N$ and an asymmetry parameter $\tau \in (0, 1)$, define the τ-*sample-expectile* $\epsilon_{\vec{x}}(\tau)$:

$$\epsilon_{\vec{x}}(\tau) = \arg\min_{\epsilon} \frac{1}{N} \sum_{i=0}^{N} [\![x_i > \epsilon]\!]_{1-\tau}^{\tau} (x_i - \epsilon)^2. \tag{3}$$

Differentiating the optimisation target of (3) with respect to ϵ yields a necessary condition for $\epsilon = \epsilon_{\vec{x}}(\tau)$:

$$0 = \frac{-2}{N} \sum_{i=0}^{N} [\![x_i > \epsilon]\!]_{1-\tau}^{\tau} (x_i - \epsilon) \qquad \text{or, equivalently,}$$

$$0 = (1 - \tau) \sum_{i:x_i \leq \epsilon} \frac{1}{N} (x_i - \epsilon) + \tau \sum_{i:x_i > \epsilon} \frac{1}{N} (x_i - \epsilon). \tag{4}$$

Define the *sample cumulative distribution function* $F_{\vec{x}}(\epsilon)$ and the *sample partial moment function* $M_{\vec{x}}(\epsilon)$, and note their 'complement' identities:

$$F_{\vec{x}}(\epsilon) = \sum_{i:x_i \leq \epsilon} \frac{1}{N} \qquad\qquad = 1 - \sum_{i:x_i > \epsilon} \frac{1}{N} \tag{5}$$

$$M_{\vec{x}}(\epsilon) = \sum_{i:x_i \leq \epsilon} \frac{1}{N} x_i \qquad\qquad = \mu_{\vec{x}} - \sum_{i:x_i > \epsilon} \frac{1}{N} x_i \tag{6}$$

$\mu_{\vec{x}} = \sum_{i=0}^{N} \frac{1}{N} x_i = M_{\vec{x}}(\max_i x_i)$ is the sample mean.

With these functions, restate the right-hand side of (4) as a function, $G_{\vec{x}}(\epsilon)$ as in (7), for which we seek a root.

$$G_{\vec{x}}(\epsilon) = -((1 - \tau) F_{\vec{x}}(\epsilon) + \tau(1 - F_{\vec{x}}(\epsilon))) \cdot \epsilon + (1 - \tau) M_{\vec{x}}(\epsilon) + \tau(\mu_{\vec{x}} - M_{\vec{x}}(\epsilon)) \tag{7}$$

Note that $F_{\vec{x}}(\epsilon)$ and $M_{\vec{x}}(\epsilon)$ are piece-wise constant (with a discontinuity at each x_i). Therefore, $G_{\vec{x}}(\epsilon)$ is piece-wise linear. Each piece has a negative slope. Moreover, $G_{\vec{x}}(\epsilon)$ is continuous[2]. This leads to the following efficient algorithm for finding the root of $G_{\vec{x}}(\epsilon)$ and thus computing the τ-sample-expectile of $\vec{x}$:

1. Sort $\vec{x}$. (Below, assume that $x_1 \leq \ldots \leq x_N$.)

2. Compute $F_{\vec{x}}(x_i)$ and $M_{\vec{x}}(x_i)$ for $i = 1, \ldots, N$. Since $\vec{x}$ is sorted, these take particularly simple forms:

$$F_i \leftarrow F_{\vec{x}}(x_i) = \frac{i}{N}, \quad M_i \leftarrow M_{\vec{x}}(x_i) = \sum_{j=1}^{i} \frac{x_j}{N}.$$

   Note that $M_N = \mu_{\vec{x}}$ and $F_N = 1$.

3. Compute $G_{\vec{x}}(x_i)$ for $i = 1, \ldots, N$:

$$G_i \leftarrow -((1-\tau)F_i + \tau(F_N - F_i)) \cdot x_i + (1-\tau)M_i + \tau(M_N - M_i).$$

4. The ith segment of $G_{\vec{x}}(\epsilon)$ runs from (x_i, G_i) down to (x_{i+1}, G_{i+1}). Find the i with $G_i \geq 0 > G_{i+1}$ and compute this segment's root:

$$\frac{(1-\tau)M_i + \tau(M_N - M_i)}{(1-\tau)F_i + \tau(F_N - F_i)}.$$

   This is both $G_{\vec{x}}(\epsilon)$'s root and the τ-sample-expectile.

Finally, note that omitting the $\frac{1}{N}$ factor in the computation of F_i and M_i may assist with numerical stability.

References

[1] Dennis J Aigner, Takeshi Amemiya, and Dale J Poirier. On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, pages 377–396, 1976.

[2] Whitney K Newey and James L Powell. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pages 819–847, 1987.

[3] Linda Schulze Waltrup, Fabian Sobotka, Thomas Kneib, and Göran Kauermann. Expectile and quantile regression—david and goliath? *Statistical Modelling*, 15(5):433–456, 2015.

---

[2]This is despite discontinuities in $F_{\vec{x}}(\epsilon)$ and $M_{\vec{x}}(\epsilon)$. One may verify that $\lim_{\epsilon \uparrow x_i} G_{\vec{x}}(\epsilon) = G_{\vec{x}}(x_i)$ for $i = 2, \ldots, N$ because $x_i F_{\vec{x}}(x_i) - x_i F_{\vec{x}}(x_{i-1}) = x_i/N = M_{\vec{x}}(x_i) - M_{\vec{x}}(x_{i-1})$. We omit a full proof.

[4] M Chris Jones. Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153, 1994.

[5] Thomas Kneib. Beyond mean regression. *Statistical Modelling*, 13(4):275–303, 2013.

[6] Bradley Efron. Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, pages 93–125, 1991.

[7] James W Taylor. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252, 2008.

[8] Tetsuro Morimura, Masashi Sugiyama, Hisashi Kashima, Hirotaka Hachiya, and Toshiyuki Tanaka. Parametric return density estimation for reinforcement learning. *arXiv preprint arXiv:1203.3497*, 2012.