Supplementary Document

This document is intended to be used as a supplement to the paper "What is the value of experimentation & measurement?" by Liu and Chamberlain. We first show in Appendix A a full derivation of the quantities presented in Section IV of the paper. This is followed by three empirical extensions to the model described in the section that open the door for future work in Appendix B.

Appendix A
Full Derivation of Valuation Under Independent Gaussian Assumptions

In this appendix we provide the full derivation of the quantities presented in Section IV of the paper. We begin by showing the standard Bayesian inference result (in one-dimensional form) quoted in Equations (6) and (7) in Section A-A. This is followed by a statement of the result in its multi-dimensional form, and the derivation of its specialisation used in Equation (15) in Section A-B. We finally provide the full derivation of that presented in Sections IV-A and IV-B in Sections A-C and A-D respectively.

A. Mean & variance of a conditioned normal r.v.

We first replicate the setup in Section IV of the paper. Let $X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ and $\epsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$, where all X_ns and ϵ_ns are mutually independent. In addition let $Y_n = X_n + \epsilon_n$, this means

$$Y_n \sim \mathcal{N}(\mu_X + \mu_\epsilon\,,\, \sigma_X^2 + \sigma_\epsilon^2) \tag{21}$$

and the conditional distribution of Y_n given $X_n = x$ is

$$(Y_n \,|\, X_n = x) \;\sim\; \mathcal{N}(x + \mu_\epsilon,\, \sigma_\epsilon^2). \tag{22}$$

A standard Bayesian inference result states the conditional distribution of X_n given $Y_n = y$ is a normal distribution with

$$\mu_{X_n|(Y_n=y)} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2}(y - \mu_\epsilon) + \frac{\sigma_\epsilon^2}{\sigma_X^2 + \sigma_\epsilon^2}\mu_X, \tag{23}$$

$$\sigma^2_{X_n|(Y_n=y)} = \frac{\sigma_\epsilon^2 \sigma_X^2}{\sigma_X^2 + \sigma_\epsilon^2}\,. \tag{24}$$

We will show below that this is indeed the case. We begin by specifying the conditional distribution $X_n \,|\, Y_n$ as the posterior distribution given X_n as the prior and $Y_n \,|\, X_n$ as the likelihood:

$$f_{X_n|Y_n}(x \,|\, y) = \frac{f_{Y_n|X_n}(y \,|\, x)\, f_{X_n}(x)}{f_{Y_n}(y)}$$
$$= \frac{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_\epsilon^2}} \exp\Big(-\frac{1}{2}\frac{(y-(x+\mu_\epsilon))^2}{\sigma_\epsilon^2}\Big) \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2}} \exp\Big(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\Big)}{\frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2+\sigma_\epsilon^2}} \exp\Big(-\frac{1}{2}\frac{(y-(\mu_X+\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\Big)}. \tag{25}$$

Grouping the fraction terms and exponential terms together and simplifying them, the RHS of Equation (25) becomes

$$\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}} \exp\Bigg(-\frac{1}{2}\Bigg[\frac{(y-(x+\mu_\epsilon))^2}{\sigma_\epsilon^2} + \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{(y-(\mu_X+\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\Bigg]\Bigg). \tag{26}$$

We then make x the principal term in the squared expressions. Note we can rewrite $(y-(x+\mu_\epsilon))^2$ as $(x-(y-\mu_\epsilon))^2$ and $(y-(\mu_X+\mu_\epsilon))^2$ as $(\mu_X-(y-\mu_\epsilon))^2$ likewise. Expression (26) can then be written as

$$\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}} \exp\Bigg(-\frac{1}{2}\Bigg[\frac{(x-(y-\mu_\epsilon))^2}{\sigma_\epsilon^2} + \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\Bigg]\Bigg). \tag{27}$$

We now attempt to complete the square. The expression within the exponent can be grouped to the x^2, x, and constant terms, giving:

$$\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}} \exp\Bigg(-\frac{1}{2}\Bigg[x^2\left[\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_X^2}\right] - 2x\left[\frac{y-\mu_\epsilon}{\sigma_\epsilon^2}+\frac{\mu_X}{\sigma_X^2}\right] + \frac{(y-\mu_\epsilon)^2}{\sigma_\epsilon^2} + \frac{\mu_X^2}{\sigma_X^2} - \frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}\Bigg]\Bigg). \tag{28}$$

Note we can rewrite the coefficient of each of the three terms with $\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}$, the intended variance of the posterior distribution, as the denominator. The coefficient for the x^2 and x terms are straightforward:

$$\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_X^2} = \frac{\sigma_X^2+\sigma_\epsilon^2}{\sigma_\epsilon^2\sigma_X^2} = \frac{1}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\,, \tag{29}$$

$$\frac{y-\mu_\epsilon}{\sigma_\epsilon^2} + \frac{\mu_X}{\sigma_X^2} = \frac{\sigma_X^2(y-\mu_\epsilon)+\sigma_\epsilon^2(\mu_X)}{\sigma_\epsilon^2\sigma_X^2} = \frac{\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon) + \frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}(\mu_X)}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\,. \tag{30}$$

To complete the square in Expression (28) properly, we expect to see the numerator for the constant term to be the square of the numerator in the RHS of Expression (30). We first transform the constant term to one that contains the desired denominator:

$$\frac{(y-\mu_\epsilon)^2}{\sigma_\epsilon^2}+\frac{\mu_X^2}{\sigma_X^2}-\frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2}$$
$$=\frac{\sigma_X^2(y-\mu_\epsilon)^2+\sigma_\epsilon^2\mu_X^2-\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(\mu_X-(y-\mu_\epsilon))^2}{\sigma_\epsilon^2\sigma_X^2}$$
$$=\frac{\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon)^2+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X^2-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}(\mu_X-(y-\mu_\epsilon))^2}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}. \tag{31}$$

This is followed by extracting the $(y-\mu_\epsilon)^2$ and μ_X^2 terms from the numerator of Expression (31), giving

$$\frac{1}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\Bigg[(y-\mu_\epsilon)^2\left[\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}\right]+ \mu_X^2\left[\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}\right]+ 2\mu_X(y-\mu_\epsilon)\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}\Bigg]. \tag{32}$$

By observing $\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}-\frac{\sigma_\epsilon^2\sigma_X^2}{(\sigma_X^2+\sigma_\epsilon^2)^2}=\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2$ and likewise for the σ_ϵ^2 case, as well as expanding the last numerator term of Expression (32), the constant term can be written as

$$\frac{1}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\Bigg[(y-\mu_\epsilon)^2\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2+\mu_X^2\left(\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2+ 2(y-\mu_\epsilon)\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\,\mu_X\,\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\Bigg], \tag{33}$$

which the numerator is simply the expected result expanded:

$$\frac{(y-\mu_\epsilon)^2}{\sigma_\epsilon^2}+\frac{\mu_X^2}{\sigma_X^2}-\frac{(\mu_X-(y-\mu_\epsilon))^2}{\sigma_X^2+\sigma_\epsilon^2} =\frac{\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}(\mu_X)\right)^2}{\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}}\ . \tag{34}$$

Substituting Equations (29), (30), and (34) into Expression (28), we arrive at the PDF of the posterior distribution:

$$f_{X_n|Y_n}(x\,|\,y)$$
$$=\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{X_n|Y_n}}}\exp\left(-\frac{1}{2}\left[\frac{x^2-2x(\mu_{X_n|Y_n})+(\mu_{X_n|Y_n})^2}{\sigma^2_{X_n|Y_n}}\right]\right)$$
$$=\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2_{X_n|Y_n}}}\exp\left(-\frac{1}{2}\left[\frac{(x-\mu_{X_n|Y_n})^2}{\sigma^2_{X_n|Y_n}}\right]\right), \tag{35}$$

where $\mu_{X_n|Y_n}$ and $\sigma^2_{X_n|Y_n}$ are that defined in Equations (23) and (24) respectively. This is clearly the PDF of a normal distribution.

B. Covariance of a conditioned normal r.v.

In this section we state the standard Bayesian inference result in the multi-dimensional case. We then use the generalized result to show that if the X_ns and Y_ns are set up as above, any X_i and X_j ($i \neq j$) will still be uncorrelated even when the values of the corresponding Y_i and Y_j are known. This result will be used when we derive Equation (15) in full in Section A-D.

The general result:

Eaton [6] has provided the following result regarding the conditional distribution for a multivariate normal distribution. Let $\boldsymbol{x}$ be a n-dimensional multivariate normal random variable. If we partition $\boldsymbol{x}$ into two components of dimension q and $n-q$, and its mean and covariance matrix accordingly such that

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}\sim\mathcal{N}\left(\boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right), \tag{36}$$

where $\boldsymbol{x}_1,\boldsymbol{\mu}_1\in\mathbb{R}^q$, $\boldsymbol{x}_1,\boldsymbol{\mu}_1\in\mathbb{R}^{n-q}$, and $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$, and $\boldsymbol{\Sigma}_{22}$ have sizes $q\times q$, $q\times(n-q)$, $(n-q)\times q$, and $(n-q)\times(n-q)$ respectively, then the distribution of $\boldsymbol{x}_1$ conditional on $\boldsymbol{x}_2=\boldsymbol{a}$ is also a multivariate normal with:

$$(\boldsymbol{x}_1|\boldsymbol{x}_2=\boldsymbol{a})$$
$$\sim\mathcal{N}\big(\boldsymbol{\mu}_1+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{a}-\boldsymbol{\mu}_2)\,,\ \boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\big). \tag{37}$$

Specializing the result to our setup:

To show the uncorrelatedness of X_ns conditional on corresponding Y_ns, we revisit the setup in Section A-A and focus on the i^{th}- and j^{th}-indexed random variables. This allows us to construct a four-dimensional vector $\boldsymbol{x}=(X_i,X_j,Y_i,Y_j)^T$ and partition them into two two-dimensional components, so that the general result can be applied.

The setup specifies that $X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X,\sigma_X^2)$, $\epsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_\epsilon,\sigma_\epsilon^2)$, and $X_i \perp \epsilon_j\ \forall\, i,j$. For the i^{th}- and j^{th}-indexed random variables we have

$$X_i\sim\mathcal{N}(\mu_X,\sigma_X^2)\,,\ X_j\sim\mathcal{N}(\mu_X,\sigma_X^2), \tag{38}$$
$$\epsilon_i\sim\mathcal{N}(\mu_\epsilon,\sigma_\epsilon^2)\,,\ \epsilon_j\sim\mathcal{N}(\mu_\epsilon,\sigma_\epsilon^2), \tag{39}$$

where all the four random variables are independent from each other. The setup also specifies that $Y_n=X_n+\epsilon_n$, this yields

$$Y_i\sim\mathcal{N}(\mu_X+\mu_\epsilon,\ \sigma_X^2+\sigma_\epsilon^2),\text{ and } Y_j\sim\mathcal{N}(\mu_X+\mu_\epsilon,\ \sigma_X^2+\sigma_\epsilon^2), \tag{40}$$

where $X_i \perp Y_j\ \forall i\neq j$.

We then construct the mean vector and covariance matrix required by the general result. Expressions (38) and (40) provided the information required to complete the entire mean vector and most of the covariance matrix. We only need to obtain $\text{Cov}(X_i,Y_i)$ (and $\text{Cov}(X_j,Y_j)$, which amounts to the same quantity). By definition of covariance:

$$\text{Cov}(X_i,Y_i)=\mathbb{E}(X_iY_i)-\mathbb{E}(X_i)\mathbb{E}(Y_i). \tag{41}$$

The first term is calculated by observing $Y_i = X_i + \epsilon_i$, $X_i \perp \epsilon_i$, and standard identities between expectations and variances:

$$\begin{aligned}\mathbb{E}(X_iY_i) &= \mathbb{E}(X_i(X_i+\epsilon_i)) = \mathbb{E}(X_i^2) + \mathbb{E}(X_i\epsilon_i)\\ &= \mathbb{E}^2(X_i) + \text{Var}(X_i) + \mathbb{E}(X_i)\mathbb{E}(\epsilon_i).\end{aligned} \tag{42}$$

Using Expressions (38) and (39) when substituting the expectations and variances, we have from the above

$$\begin{aligned}&\text{Cov}(X_i, Y_i)\\ &= \mathbb{E}^2(X_i) + \text{Var}(X_i) + \mathbb{E}(X_i)\mathbb{E}(\epsilon_i) - \mathbb{E}(X_i)\mathbb{E}(Y_i)\\ &= \mu_X^2 + \sigma_X^2 + \mu_X\mu_\epsilon - \mu_X(\mu_X + \mu_\epsilon) = \sigma_X^2.\end{aligned} \tag{43}$$

This yields the mean vector and covariance matrix for $\boldsymbol{x} = (X_i, X_j, Y_i, Y_j)^T$ as

$$\boldsymbol{\mu} = \begin{pmatrix}\mu_X\\ \mu_X\\ \mu_X+\mu_\epsilon\\ \mu_X+\mu_\epsilon\end{pmatrix},\ \boldsymbol{\Sigma} = \begin{pmatrix}\sigma_X^2 & 0 & \sigma_X^2 & 0\\ 0 & \sigma_X^2 & 0 & \sigma_X^2\\ \sigma_X^2 & 0 & \sigma_X^2+\sigma_\epsilon^2 & 0\\ 0 & \sigma_X^2 & 0 & \sigma_X^2+\sigma_\epsilon^2\end{pmatrix}. \tag{44}$$

Partitioning $\boldsymbol{x}$ into two components $(X_i, X_j)^T$ and $(Y_i, Y_j)^T$, and using the general result above we then have:

$$\begin{pmatrix}X_i\\ X_j\end{pmatrix} \Bigg| \begin{pmatrix}Y_i\\ Y_j\end{pmatrix} = \begin{pmatrix}y_i\\ y_j\end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Sigma}') \tag{45}$$

where

$$\begin{aligned}\boldsymbol{\mu}' = &\begin{pmatrix}\mu_X\\ \mu_X\end{pmatrix} + \begin{pmatrix}\sigma_X^2 & 0\\ 0 & \sigma_X^2\end{pmatrix}\cdot\\ &\begin{pmatrix}\sigma_X^2+\sigma_\epsilon^2 & 0\\ 0 & \sigma_X^2+\sigma_\epsilon^2\end{pmatrix}^{-1}\left[\begin{pmatrix}y_i\\ y_j\end{pmatrix} - \begin{pmatrix}\mu_X+\mu_\epsilon\\ \mu_X+\mu_\epsilon\end{pmatrix}\right],\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}' = &\begin{pmatrix}\sigma_X^2 & 0\\ 0 & \sigma_X^2\end{pmatrix} - \begin{pmatrix}\sigma_X^2 & 0\\ 0 & \sigma_X^2\end{pmatrix}\cdot\\ &\begin{pmatrix}\sigma_X^2+\sigma_\epsilon^2 & 0\\ 0 & \sigma_X^2+\sigma_\epsilon^2\end{pmatrix}^{-1}\begin{pmatrix}\sigma_X^2 & 0\\ 0 & \sigma_X^2\end{pmatrix}.\end{aligned} \tag{46}$$

Simplifying the expression by standard matrix operations we arrive at:

$$\boldsymbol{\mu}' = \begin{pmatrix}\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X + \frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y_i - \mu_\epsilon)\\ \frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X + \frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y_j - \mu_\epsilon)\end{pmatrix}, \tag{47}$$

$$\boldsymbol{\Sigma}' = \begin{pmatrix}\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2} & 0\\ 0 & \frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\end{pmatrix}. \tag{48}$$

Note the mean and variance of X_i and X_j is that derived in Section A-A. Furthermore, the covariance between X_i and X_j given Y_i and Y_j is zero, as claimed at the beginning of this subsection.

C. Calculating the Expectation

The results above are necessary to derive the quantities presented in Section IV of the paper. Here we go through the derivation in greater detail, focusing more on the algebraic manipulations and being light on the commentaries found in the main paper. Readers are encouraged to read Section III of the paper to familiarize themselves with the notations and terminologies used below before proceeding.

We are interested in deriving the expected value of D, the value gained when the estimation noise is reduced. To do so we require the expected values of, in order:

1) $Y_{(r)}$ - the *estimated* value of the r^{th} proposition, ranked in increasing estimated value;
2) $X_{\mathcal{I}(r)}$ - the *true* value of the r^{th} proposition, ranked by increasing estimated value
3) V - the mean of the *true* value for the M most valuable propositions, ranked by their estimated values.

To obtain the expected value for $Y_{(r)}$, we begin by observing that the Y_n are normally distributed. Using the standard properties of the normal distribution:

$$Y_n = X_n + \epsilon_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_X + \mu_\epsilon,\, \sigma_X^2 + \sigma_\epsilon^2). \tag{49}$$

This is followed by applying a result by Blom [1], which states that the expected value for normal order statistics $Y_{(r)}$ can be closely approximated as:

$$\mathbb{E}(Y_{(r)}) \approx \mu_X + \mu_\epsilon + \sqrt{\sigma_X^2 + \sigma_\epsilon^2}\ \Phi^{-1}\left(\frac{r-\alpha}{N-2\alpha+1}\right), \tag{50}$$

where Φ^{-1} denotes the quantile function of a standard normal distribution, and α is a constant.[10]

The expected value of $X_{\mathcal{I}(r)}$ is obtained as follows. We first recall two standard results in Bayesian inference, the first being the conditional probability distribution for any Y_n given X_n is normally distributed:

$$Y_n|(X_n = x)\ \sim\ \mathcal{N}(x + \mu_\epsilon, \sigma_\epsilon^2). \tag{51}$$

The second result states that the posterior distribution of X_n once Y_n is observed is also normally distributed, with mean $\mu_{X_n|Y_n}$ and variance $\sigma^2_{X_n|Y_n}$ given by (see Section A-A):

$$\mu_{X_n|(Y_n=y)} = \frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y - \mu_\epsilon) + \frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X, \tag{52}$$

$$\sigma^2_{X_n|Y_n} = \frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}. \tag{53}$$

We then apply the law of iterated expectations to obtain

$$\mathbb{E}(X_{\mathcal{I}(r)}) = \mathbb{E}_{Y_{(r)}}\left(\mathbb{E}(X_{\mathcal{I}(r)}\,|\,Y_{(r)})\right). \tag{54}$$

Noting $Y_{\mathcal{I}(r)}$ is equivalent to $Y_{(r)}$ as the propositions are ranked by their estimated values, we substitute Equation (52)

[10] While there are a number of literature on the best value of α, the exact value of this constant does not play a huge role as long as $\alpha \approx 0.4$ [7].

into Equation (54), and move the constant terms out of the outer expectation to get

$$\mathbb{E}(X_{\mathcal{I}(r)}) = \mathbb{E}_{Y_{(r)}}\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X\right)$$
$$= \frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\mathbb{E}_{Y_{(r)}}(y) - \frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_\epsilon + \frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X . \tag{55}$$

Further substituting Equation (50) into Equation (55) and simplifying the equation yields

$$\begin{aligned}&\mathbb{E}(X_{\mathcal{I}(r)})\\ &\approx \frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\left(\mu_X+\mu_\epsilon+\sqrt{\sigma_X^2+\sigma_\epsilon^2}\,\Phi^{-1}\left(\frac{r-\alpha}{N-2\alpha+1}\right)\right)\\ &\quad-\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_\epsilon+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X .\end{aligned} \tag{56}$$

The μ_ϵ terms cancel, and the μ_X terms sum to μ_X leading to

$$\begin{aligned}\mathbb{E}(X_{\mathcal{I}(r)}) &\approx \mu_X + \frac{\sigma_X^2\sqrt{\sigma_X^2+\sigma_\epsilon^2}}{\sigma_X^2+\sigma_\epsilon^2}\,\Phi^{-1}\left(\frac{r-\alpha}{N-2\alpha+1}\right)\\ &= \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_\epsilon^2}}\,\Phi^{-1}\left(\frac{r-\alpha}{N-2\alpha+1}\right).\end{aligned} \tag{57}$$

Equation (57) shows that decreasing the estimation noise σ_ϵ^2 will lead to an increase in $\mathbb{E}(X_{\mathcal{I}(r)})$ for any $r > \frac{N+1}{2}$. It follows that the mean true value of the top M propositions, selected according to their estimated value, will increase with the presence of a lower estimation noise. We show this by applying the expectation function to V defined in Equation (2):

$$\begin{aligned}\mathbb{E}(V) = \frac{1}{M}\big(&\mathbb{E}(X_{\mathcal{I}(N-M+1)})+\mathbb{E}(X_{\mathcal{I}(N-M+2)})+\\ &\cdots+\mathbb{E}(X_{\mathcal{I}(N)})\big).\end{aligned} \tag{58}$$

Substituting Equation (57) into the above gives

$$\begin{aligned}&\mathbb{E}(V)\\ &\approx \frac{1}{M}\Bigg(\mu_X+\frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_\epsilon^2}}\,\Phi^{-1}\left(\frac{N-M+1-\alpha}{N-2\alpha+1}\right)+\\ &\qquad \mu_X+\frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_\epsilon^2}}\,\Phi^{-1}\left(\frac{N-M+2-\alpha}{N-2\alpha+1}\right)+\cdots+\\ &\qquad \mu_X+\frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_\epsilon^2}}\,\Phi^{-1}\left(\frac{N-\alpha}{N-2\alpha+1}\right)\Bigg).\end{aligned} \tag{59}$$

Observing there are M copies of μ_X, and the Φ^{-1} terms can be written as a summation, we arrive at

$$\mathbb{E}(V) \approx \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_\epsilon^2}}\frac{1}{M}\sum_{r=N-M+1}^{N}\Phi^{-1}\left(\frac{r-\alpha}{N-2\alpha+1}\right). \tag{60}$$

We finally consider the improvement when we reduce the estimation noise from $\sigma_\epsilon^2=\sigma_1^2$ to σ_2^2:

$$\begin{aligned}\mathbb{E}(D) &= \mathbb{E}(V|_{\sigma_\epsilon^2=\sigma_2^2}) - \mathbb{E}(V|_{\sigma_\epsilon^2=\sigma_1^2})\\ &\approx \left(\frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_2^2}}-\frac{\sigma_X^2}{\sqrt{\sigma_X^2+\sigma_1^2}}\right)\times\\ &\quad \frac{1}{M}\sum_{r=N-M+1}^{N}\Phi^{-1}\left(\frac{r-\alpha}{N-2\alpha+1}\right).\end{aligned} \tag{61}$$

If we assume $\mu_X = 0$ (i.e. the true value of the propositions are centred around zero), then the relative gain is entirely dependent on σ_X^2, σ_1^2, σ_2^2:

$$\frac{\mathbb{E}(D|_{\mu_X=0})}{\mathbb{E}(V|_{\sigma_\epsilon^2=\sigma_1^2,\,\mu_X=0})} = \frac{\sqrt{\sigma_X^2+\sigma_1^2}}{\sqrt{\sigma_X^2+\sigma_2^2}} - 1\,. \tag{62}$$

D. Calculating the Variance

Having derived the expected value in Equations (61) and (62), in this section we address the variance of D. Deriving the variance is similar to deriving the expectation — one has to obtain the variances for (in order) $Y_{(r)}$, $X_{\mathcal{I}(r)}$, and V. For the variance of $Y_{(r)}$, we apply a result from David and Johnson [4], which states that given Y_n as defined in Expression (49), $\mathrm{Var}\left(Y_{(r)}\right)$ can be approximated as:

$$\mathrm{Var}(Y_{(r)}) \approx \frac{r(N-r+1)}{(N+1)^2(N+2)}\frac{\sigma_X^2+\sigma_\epsilon^2}{\left(\phi\left(\Phi^{-1}\left(\frac{r}{N+1}\right)\right)\right)^2}\,, \tag{63}$$

where ϕ is the probability density function, and Φ^{-1} is the quantile function of a standard normal distribution.

The variance for $X_{\mathcal{I}(r)}$ is then obtained using the law of total variance:

$$\begin{aligned}&\mathrm{Var}(X_{\mathcal{I}(r)})\\ &= \mathbb{E}_{Y_{(r)}}\left(\mathrm{Var}(X_{\mathcal{I}(r)}|Y_{(r)})\right) + \mathrm{Var}_{Y_{(r)}}\left(\mathbb{E}(X_{\mathcal{I}(r)}|Y_{(r)})\right).\end{aligned} \tag{64}$$

Recognizing $Y_{(r)}$ and $Y_{\mathcal{I}(r)}$ are equivalent, we substitute the conditional expectations stated in Equations (52) and (53) into Equation (64), and move the constant terms out of the outer expectation / variance to get

$$\begin{aligned}\mathrm{Var}(X_{\mathcal{I}(r)}) &= \mathbb{E}_{Y_{(r)}}\left(\frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)+\\ &\quad \mathrm{Var}_{Y_{(r)}}\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(y-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X\right)\\ &= \frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}+\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2\mathrm{Var}_{Y_{(r)}}(y)\,.\end{aligned} \tag{65}$$

Further substituting Expression (63) into Equation (65) and cancelling out equal terms, we have

$$\begin{aligned}\mathrm{Var}(X_{\mathcal{I}(r)}) &\approx \frac{\sigma_\epsilon^2\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}+\\ &\quad \frac{\sigma_X^4}{\sigma_X^2+\sigma_\epsilon^2}\frac{r(N-r+1)}{(N+1)^2(N+2)}\frac{1}{\left(\phi\left(\Phi^{-1}(\frac{r}{N+1})\right)\right)^2}\,.\end{aligned} \tag{66}$$

Before we derive the variance of V, we require the covariance between pairs of $Y_{(\cdot)}$s and $X_{\mathcal{I}(\cdot)}$s. This is necessary as the terms of V (see Equation (2)), being the result of removing noise from successive order statistics, are highly correlated.

David and Nagaraja [5] have provided a formula to estimate the covariance between $Y_{(r)}$ and $Y_{(s)}$ for any $r, s \leq N$, first presented by David and Johnson [4]:

$$\begin{aligned}&\mathrm{Cov}(Y_{(r)}, Y_{(s)})\\ &\approx \frac{r(N-s+1)}{(N+1)^2(N+2)} \frac{\sigma_X^2+\sigma_\epsilon^2}{\phi\left(\Phi^{-1}(\frac{r}{N+1})\right)\,\phi\left(\Phi^{-1}(\frac{s}{N+1})\right)}\,. \end{aligned} \tag{67}$$

To obtain the covariance between $X_{\mathcal{I}(r)}$ and $X_{\mathcal{I}(s)}$ for any $r, s \leq N$, we first observe the law of total covariance with multiple conditioning variables [2] states that for $r \neq s$:

$$\begin{aligned}&\mathrm{Cov}(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)})\\ =&\,\mathbb{E}\left(\mathbb{E}(\mathrm{Cov}(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)}\,|\,Y_{(r)}, Y_{(s)})\,|\,Y_{(r)})\right)+\\ &\mathbb{E}\big(\mathrm{Cov}(\mathbb{E}(X_{\mathcal{I}(r)}\,|\,Y_{(r)}, Y_{(s)}),\\ &\qquad\mathbb{E}(X_{\mathcal{I}(s)}\,|\,Y_{(r)}, Y_{(s)})\,|\,Y_{(r)})\big)+\\ &\mathrm{Cov}\big(\mathbb{E}(\mathbb{E}(X_{\mathcal{I}(r)}\,|\,Y_{(r)}, Y_{(s)})\,|\,Y_{(r)}),\\ &\qquad\mathbb{E}(\mathbb{E}(X_{\mathcal{I}(s)}\,|\,Y_{(r)}, Y_{(s)})\,|\,Y_{(r)})\big)\,. \end{aligned} \tag{68}$$

Noting that $Y_{(\cdot)}$ is equivalent to $Y_{\mathcal{I}(\cdot)}$ by definition, we observe the first term on the RHS of Equation (68) is zero. This follows from properties of a multivariate normal's conditional distributions — X_ns are uncorrelated to each other given the corresponding Y_ns if X_n themselves are uncorrelated (see Section A-B). For the second and third term, we note X_n is independent of Y_m for any $n \neq m$ and hence $\mathbb{E}(X_n|Y_n, Y_m) = \mathbb{E}(X_n|Y_n)$. This allows us to substitute Equation (52) into Equation (68) to get

$$\begin{aligned}&\mathrm{Cov}(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)}) = 0+\\ &\mathbb{E}\Bigg(\mathrm{Cov}\bigg(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(Y_{(r)}-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X,\\ &\qquad\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(Y_{(s)}-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X\,\bigg|\,Y_{(r)}\bigg)\Bigg)+\\ &\mathrm{Cov}\Bigg(\mathbb{E}\bigg(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(Y_{(r)}-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X\,\bigg|\,Y_{(r)}\bigg),\\ &\quad\mathbb{E}\bigg(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}(Y_{(s)}-\mu_\epsilon)+\frac{\sigma_\epsilon^2}{\sigma_X^2+\sigma_\epsilon^2}\mu_X\,\bigg|\,Y_{(r)}\bigg)\Bigg), \end{aligned} \tag{69}$$

which can be simplified using standard properties of expectation and covariance functions to give

$$\begin{aligned}&\mathrm{Cov}(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)})\\ =&\,\mathbb{E}\left(\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2\mathrm{Cov}\left(Y_{(r)}, Y_{(s)}\,|\,Y_{(r)}\right)\right)+\\ &\left(\frac{\sigma_X^2}{\sigma_X^2+\sigma_\epsilon^2}\right)^2\mathrm{Cov}\left(Y_{(r)}, Y_{(s)}\right)\,. \end{aligned} \tag{70}$$

Clearly $\mathrm{Cov}(A, B\,|\,A) = 0\ \forall A, B$. Hence we substitute Equation (67) into Equation (70) and simplify the resultant expression to arrive at

$$\begin{aligned}&\mathrm{Cov}(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)}) \approx\\ &\frac{\sigma_X^4}{\sigma_X^2+\sigma_\epsilon^2}\frac{r(N-s+1)}{(N+1)^2(N+2)}\frac{1}{\phi\left(\Phi^{-1}(\frac{r}{N+1})\right)\,\phi\left(\Phi^{-1}(\frac{s}{N+1})\right)}. \end{aligned} \tag{71}$$

Equation (71) affirms the claim that the $X_{\mathcal{I}(\cdot)}$s are positively correlated. Now we can state the variance of V and D. Applying the variance function to the defintion of V (Equation (2)) we get

$$\begin{aligned}&\mathrm{Var}(V)\\ &=\frac{1}{M^2}\mathrm{Var}\left(X_{\mathcal{I}(N-M+1)}+X_{\mathcal{I}(N-M+2)}+...+X_{\mathcal{I}(N)}\right)\\ &=\frac{1}{M^2}\Bigg(\sum_{r=N-M+1}^{N}\mathrm{Var}\left(X_{\mathcal{I}(r)}\right)+\\ &\qquad\sum_{r=N-M+1}^{N}\sum_{s=r+1}^{N}2\cdot\mathrm{Cov}\left(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)}\right)\Bigg), \end{aligned} \tag{72}$$

where $\mathrm{Var}(X_{\mathcal{I}(r)})$ and $\mathrm{Cov}(X_{\mathcal{I}(r)}, X_{\mathcal{I}(s)})$ are defined in Equations (66) and (71).

The variance of D is thus:

$$\begin{aligned}\mathrm{Var}(D) =&\mathrm{Var}(V|_{\sigma_\epsilon^2=\sigma_2^2}-V|_{\sigma_\epsilon^2=\sigma_1^2})\\ =&\mathrm{Var}\left(V|_{\sigma_\epsilon^2=\sigma_2^2}\right)+\mathrm{Var}\left(V|_{\sigma_\epsilon^2=\sigma_1^2}\right)\\ &-2\cdot\mathrm{Cov}\left(V|_{\sigma_\epsilon^2=\sigma_2^2}, V|_{\sigma_\epsilon^2=\sigma_1^2}\right). \end{aligned} \tag{73}$$

The first two terms on the right hand side of Equation (73) are that defined in Equation (72) (we omit the expanded form for brevity), while the last term can be expanded as follow:

$$\begin{aligned}&\mathrm{Cov}\left(V|_{\sigma_\epsilon^2=\sigma_2^2}, V|_{\sigma_\epsilon^2=\sigma_1^2}\right)\\ =&\,\mathrm{Cov}\Big(\frac{1}{M}\left(X_{\mathcal{I}(N-M+1)}+X_{\mathcal{I}(N-M+2)}+...+X_{\mathcal{I}(N)}\right)|_{\sigma_\epsilon^2=\sigma_2^2},\\ &\quad\frac{1}{M}\left(X_{\mathcal{I}(N-M+1)}+X_{\mathcal{I}(N-M+2)}+...+X_{\mathcal{I}(N)}\right)|_{\sigma_\epsilon^2=\sigma_1^2}\Big)\\ =&\,\frac{1}{M^2}\sum_{r=N-M+1}^{N}\sum_{s=r}^{N}2\cdot\mathrm{Cov}\left(X_{\mathcal{I}(r)}|_{\sigma_\epsilon^2=\sigma_2^2},\,X_{\mathcal{I}(s)}|_{\sigma_\epsilon^2=\sigma_1^2}\right)\,. \end{aligned} \tag{74}$$

Equation (74) shows the covariance term in Equation (73) is positive as all its components are positive (cf. Equation (71), albeit with a different magnitude). Hence the variance terms in Equation (73) form an upper bound to the variance of D:

$$\mathrm{Var}(D) < \mathrm{Var}\left(V|_{\sigma_\epsilon^2=\sigma_2^2}\right)+\mathrm{Var}\left(V|_{\sigma_\epsilon^2=\sigma_1^2}\right). \tag{75}$$

In practice, the variance of D is much lower than the bound, due to the Vs being highly correlated.

APPENDIX B
EMPIRICAL EXTENSIONS

We also provide three extensions, all evaluated empirically, that open the door for future work in this area.

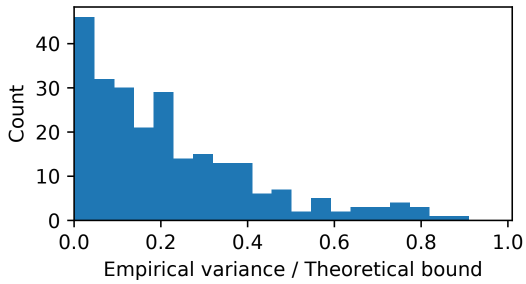


Fig. 4: The distribution of the ratios between empirical bootstrap variance estimate and the theoretical upper bound.

A. Empirical calculation of the risk

In Section IV-B (and Appendix A-D in more detail) we derived an upper bound on $\mathrm{Var}(D)$. To help understand the risk in acquiring E&M capabilities, here we perform an empirical calculation on $\mathrm{Var}(D)$ to determine how far we are from the bound in general.

Similar to the previous experiment, for each run we randomly sample N, M, μ_X, μ_ϵ, σ_X^2, σ_1^2, σ_2^2, where the restrictions $M < N$ and $\sigma_2^2 < \sigma_1^2$ are maintained, and perform 1,000 cycles of the six steps mentioned above to obtain samples of $\mathtt{D}$. This is followed by 500 bootstrap resamplings on the samples to obtain an empirical bootstrap distribution for the variance. We then take the mean of the bootstrap variance estimates, and compare them with the theoretical upper bound by computing the ratio between the empirical variance and the bound.

We performed 250 runs and show the distribution of ratios between the empirical variance and the bound in Figure 4. The figure shows while most empirical values of $\mathrm{Var}(D)$ are much lower than the upper bound, with around 85% of the samples having a ratio $\leq 40\%$, in a few scenarios, the bootstrap variance estimate is up to 90% of the upper bound, indicating a lower magnitude of the covariance term in (73).

In future work it would be interesting to establish a tighter bound for, or an accurate estimate of the variance of the value gained from prioritization under different parameter combinations.

B. Valuation under independent t-distributed assumptions

The model described in Section IV assumes that both the true value of the propositions and the E&M noise are normally distributed. While possessing decent mathematical properties, it is insufficient to explain the heavy tail in the distribution of uplifts shown in [3] or [8].

In this section, we model the true value of the propositions, as well as the estimation noise, as Generalized Student's t-distributions:[11]

$$X_n \overset{\text{i.i.d}}{\sim} t_\nu(\mu_X, \sigma_X^2),\ \epsilon_n \overset{\text{i.i.d}}{\sim} t_\nu(\mu_\epsilon, \sigma_\epsilon^2), \tag{76}$$

[11] A generalized Student's t-distribution is specified as $X = \mu + \sigma T_\nu$, where T_ν is a standard Student's t-distribution with ν degrees of freedom, and μ and σ are the location and scaling parameter respectively. The idea is similar to 'generalizing' a standard normal distribution by multiplying it with a scaling parameter and adding a location parameter.

where $\epsilon_n \perp X_m \, \forall n, m$, ν denotes the degrees of freedom of the underlying standard Student's t-distribution, μs denotes the location parameter, and σ^2s denotes the scale parameter.

It is difficult to derive theoretical quantities under such model assumptions because Student's t-distributions do not have conjugate priors (see e.g. [9]). We instead simulate the empirical distribution of the value gained under different parameter combinations to understand if this model is a better alternative to that under normal assumptions. The sampling procedure is similar to that described Section V, with Steps 1 and 2 modified such that the samples are generated from standard t-distributions, then scaled and located as specified by Expression (76).

We compare the value gain estimates obtained under t-distributed assumptions and normal assumptions as follows. For each run, we randomly sample values for N, M, μ_X, μ_ϵ, σ_X^2, σ_1^2, σ_2^2, and perform 1,000 cycles of the six-step sampling procedure Section V above to obtain samples of $\mathtt{D}$ using both the t_3 and normal distributions.[12] We then compare the expected values, as well as the 5% and 95% percentiles of the value gained under the two distributions.

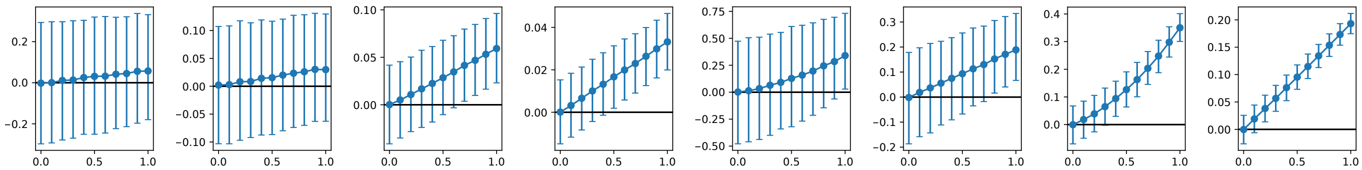
We observed from 840 runs that the distribution of value gained under the t-distributed assumptions has a higher mean (30% higher on average) and variance (40% higher in the 95% percentile on average) than under normal assumptions, reflecting the larger spread in true value and estimation noise. Moreover, if we scale the initial t-distributions by $\sqrt{(\nu-2)/\nu}$ such that it has the same variance as the normal distributions, the observation still holds, albeit with a lower magnitude (7% higher mean and 7% higher 95% percentile on average). This shows that the model under t-distributed assumptions is able to capture the "higher risk, higher reward" concept.

C. Partial estimation / measurement noise reduction

There are many situations when not all propositions are immediately measurable upon the acquisition of E&M capabilities. This may be due to the extra work required to integrate additional capabilities in certain legacy systems, or the limited ability to run experiments on online but not offline activities. In the case where there is a single backlog, we ask the question, will an organization still benefit from a partial noise reduction when some propositions' values are obtained under reduced uncertainty while others are subject to the original noise level?

We address this by attempting to establish the relationship between the expected improvement in mean true value of the selected propositions and the proportion of propositions that benefited from a reduced estimation noise (denoted $p \in [0, 1]$). The sampling procedure is similar to that described in Section V, with Step 5 modified: when we repeat Step 2, instead of generating all samples from $\mathcal{N}(\mu_\epsilon, \sigma_2^2)$ we generate p of the samples from $\mathcal{N}(\mu_\epsilon, \sigma_2^2)$ (the lowered estimation / measurement noise) and $1-p$ of the samples from $\mathcal{N}(\mu_\epsilon, \sigma_1^2)$ (the original estimation / measurement noise).

[12] t_3 (t-distribution with three degrees of freedom (d.f.)) is used as it is the distribution with the longest tail under the t family with a natural number d.f. while retaining a finite variance.



(a) $N = 50$, $M = 5$ (b) $N = 50$, $M = 20$ (c) $N = 2500$, $M = 250$ (d) $N = 2500$, $M = 1000$ (e) $N = 50$, $M = 5$ (f) $N = 50$, $M = 20$ (g) $N = 2500$, $M = 250$. (h) $N = 2500$, $M = 1000$

Fig. 5: The near-linear relationship between p (proportion of propositions which value is obtained under a lower estimation / measurement noise, x-axes) and the improvement in mean true value of the selected propositions (y-axes) under the normal value / normal noise model. In each plot the dot represents the sample mean, and the error bar represents the 5% and 95% percentile of the sample value gained (see Equation (61)). All figures assume $\sigma_X^2 = 1$, while the left four figures assume $\sigma_1^2 = 0.5^2$ and $\sigma_2^2 = 0.4^2$ (corresponding to a small reduction in estimation / measurement noise), and the right four figures assume $\sigma_1^2 = 0.8^2$ and $\sigma_2^2 = 0.2^2$ (corresponding to a large reduction in estimation / measurement noise).

We run the procedure above under various scenarios, including under a large/small (N), a large/small ratio between an organizations' capacity and backlog (M/N), and a large/small magnitude of noise reduction upon acquisition of E&M capabilities ($\sigma_1^2 - \sigma_2^2$). Figure 5 shows the result. We can see that under most scenarios, the expected value gained increases with p at least linearly, while there are a few scenarios where the expected improvement in mean true value of the selected propositions curve upwards for increasing p. This shows that while there are incentives for organizations to acquire E&M capabilities that cover the majority of their work, in many scenarios, a partial acquisition yields proportional benefits. Potential experimenters need not see the acquisition as a zero-one decision, or worry about any steep initial investment required to unlock returns.

References

[1] Blom, G. (1958). *Statistical Estimated and Transformed Beta-variables.* PhD thesis, Stockholm College.

[2] Bowsher, C. G. and Swain, P. S. (2012). "Identifying Sources of Variation and the Flow of Information in Biochemical Networks" *Proceedings of the National Academy of Sciences*, vol. 109, no. 20, pp. E1320–E1328.

[3] Browne, W. and Jones, M. S. (2017). "What Works In e-Commerce — A Meta-analysis of 6700 Online Experiments." https://www.qubit.com/wp-content/uploads/2017/12/qubit-research-meta-analysis.pdf.

[4] David, F. N. and Johnson, N. L. (1954). "Statistical Treatment of Censored Data: Part I. Fundamental Formulae." *Biometrika*, vol. 41, pp. 228–240.

[5] David, H. A. and Nagaraja, H. N. (2004). "Order Statistics." *Encyclopedia of Statistical Sciences.*

[6] Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* John Wiley and Sons. pp. 116–117. ISBN 978-0-471-02776-8.

[7] Harter, H. L. (1961). "Expected Values of Normal Order Statistics". *Biometrika*, vol. 48, pp. 151–165.

[8] Johnson, G., Lewis, R. A. and Nubbemeyer, E. (2017). "The Online Display Ad Effectiveness Funnel & Carryover: Lessons From 432 Field Experiments." http://dx.doi.org/10.2139/ssrn.2701578.

[9] Robert, C. P. (2007). *The Bayesian Choice: From Decision-theoretic Foundations to Computational Implementation.* Springer Science & Business Media.