On Information–Estimation Relationships over Binomial and Negative Binomial Models

Dongning Guo
Department of Electrical Engineering and Computer Science
Northwestern University
Evanston, IL 60208, USA

***Abstract*—In recent years, a number of results have been developed which connect information measures and estimation measures under various models, including, predominantly, Gaussian and Poisson models. More recent results due to Gil Taborda and Pérez-Cruz relate the relative entropy to certain mismatched estimation errors in the context of binomial and negative binomial models, where, unlike in the case of Gaussian and Poisson models, the conditional mean estimates concern models of different orders than those of the original model. In this paper, a different set of results in simple forms are developed for binomial and negative binomial models, where the conditional mean estimates are produced through the original models. The new results are consistent with previous results for Gaussian and Poisson models.**

I. Introduction

Since a simple differential relationship between the mutual information and the minimum mean square error over a scalar Gaussian model was discovered [1], a number of similar results have been developed for several other models, e.g., Poisson models [2], [3]. In the context of Gaussian and Poisson models, it has also been found that the relative entropy can be expressed as the integral of the increase of certain estimation errors due to mismatch in the prior distribution [3]–[5].

More recently, Gil Taborda and Pérez-Cruz [6] developed results in the context of binomial and negative binomial models. The key result expresses the derivative of the relative entropy of two output distributions of the same binomial (or negative binomial) model induced by two different inputs in terms of the penalty in a certain estimation error due to mismatch in the prior distribution. As in the case of Gaussian and Poisson models, the errors concern conditional mean estimates of the input given the output. However, the conditional mean is produced using binomial (or negative binomial) models with the order lowered or raised by one. In particular, in case of the binomial model, the modified model is of one fewer trial than the original model; in case of the negative binomial model, the modified model is of one more failure than the original model.

In this paper, we develop a different set of results concerning essentially the same binomial and negative binomial models as in [6]. The results put the derivative of the relative entropy in a simple form concerning some average difference of conditional mean estimates due to mismatched prior distribution. In contrast to the results of [6], the conditional mean estimates here are based on the original binomial and negative binomial models. The results are thus consistent with previous results for Gaussian and Poisson models [7].

This work was supported in part by the National Science Foundation under Grant No. 1231828.

II. The Binomial Model

The binomial model is based on the binomial distribution, Binomial(n, q), which describes the probability of having k successful trials in n independent Bernoulli trials, each with probability q to succeed:

$$\mathsf{P}\left(Y = k\right) = \binom{n}{k} q^k (1-q)^{n-k}, \;\; k = 0, \dots, n. \tag{1}$$

Let a be a positive parameter which takes its values in $(0, a_{\max})$. With some hindsight, we define the binomial model to be a random transformation from a random variable X which takes its values in $(a_{\max}, \infty)$ to another random variable Y, where, conditioned on $X = x$, Y follows the distribution of Binomial$(n, a/x)$. The conditional probability mass function (pmf) is given by

$$P^n_{Y|X}(y|x) = \binom{n}{y} \left(\frac{a}{x}\right)^y \left(1 - \frac{a}{x}\right)^{n-y}, \; y = 0, \dots, n \tag{2}$$

where its dependence on parameter n is made explicit. The variables X and Y are viewed as the input and output of the binomial model. Here the input X controls the probability of success of an individual Bernoulli trial, namely, for fixed X, the success-to-failure ratio is a to X. The larger X is, the fewer trials succeed on average. The parameter a is viewed as a scaling of the input.

If the prior distribution of the input X is P_X, the corresponding output distribution is denoted by P^n_Y; if the prior distribution of X is Q_X, the corresponding output distribution is denoted by Q^n_Y. The dependence of Y on the parameter a is implicit for notational convenience. Throughout this paper, we use $\mathsf{E}^n\left\{\cdot\right\}$ and $\mathsf{E}^n\left\{\cdot|\cdot\right\}$ to denote expectation and conditional expectation under distribution $P^n_{XY} = P_X P^n_{Y|X}$, whereas we use $\mathsf{E}^n_Q\left\{\cdot\right\}$ and $\mathsf{E}^n_Q\left\{\cdot|\cdot\right\}$ to denote expectation and conditional expectation under distribution $Q^n_{XY} = Q_X P^n_{Y|X}$. In particular, the conditional mean of X given Y is denoted by $\mathsf{E}^n\left\{X|Y\right\}$ under distribution P^n and by $\mathsf{E}^n_Q\left\{X|Y\right\}$ under distribution Q^n. The superscript n is often omitted when no ambiguity arises.

We also define the following function

$$g(t) = t - 1 - \log t. \tag{3}$$

This function is convex on $(0, \infty)$, and achieves its unique minimum, 0, at $t = 1$. For two positive numbers, x and $\hat{x}$, the following function of two arguments,

$$g\left(\frac{x}{\hat{x}}\right) = \frac{x}{\hat{x}} - 1 - \log\left(\frac{x}{\hat{x}}\right) \tag{4}$$

is a Bregman divergence [8], also known as the Itakura–Saito distance. The function is always nonnegative, and that it is equal to 0 if and only if $x = \hat{x}$. Moreover, $g(x/\hat{x})$ increases monotonically as $x/\hat{x}$ departs from 1 in either direction of the axis.

Proposition 1. *Let X be an arbitrary random variable which takes its values in $(a_{\max}, \infty)$. Conditioned on X, let Y be a binomial random variable with parameters $(n, a/X)$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}a} I(X;Y) = \mathsf{E}\left\{\frac{Y}{a} \cdot g\left(\frac{X - a}{\mathsf{E}\{X|\, Y\} - a}\right)\right\} \tag{5}$$

holds for all $a \in (0, a_{\max})$.

Proposition 2. *Let P_Y and Q_Y be the output distribution of the binomial model* (2) *induced by input distributions P_X and Q_X, respectively, where P_X and Q_X put no probability mass on $(-\infty, a_{\max}]$. Then for every $a \in (0, a_{\max})$,*

$$\frac{\mathrm{d}}{\mathrm{d}a} \mathsf{D}\left(P_Y \| Q_Y\right) = \mathsf{E}\left\{\frac{Y}{a} \cdot g\left(\frac{\mathsf{E}\{X|\, Y\} - a}{\mathsf{E}_Q\{X \mid Y\} - a}\right)\right\}. \tag{6}$$

Before we prove Propositions 1 and 2, it is interesting to compare them with the following result:[1]

Proposition 3 (Gil Taborda and Pérez-Cruz [6])**.** *Let P, Q, X and Y be defined as in Proposition 2. Then, for every $a \in (0, a_{\max})$,*

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}a} D(P_Y^n \| Q_Y^n) \\ &= \frac{n}{a} \mathsf{E}_P^{n-1}\left\{\ell(\mathsf{E}_P^{n-1}\left\{aX^{-1}\middle| Y\right\}, \mathsf{E}_Q^{n-1}\left\{aX^{-1}\middle| Y\right\})\right\} \end{aligned} \tag{7}$$

where the right hand side involves a Bregman divergence:

$$\ell(a, \hat{a}) = a \log \frac{a(1 - \hat{a})}{\hat{a}(1 - a)} - \frac{a - \hat{a}}{1 - \hat{a}}. \tag{8}$$

The distributions and expectations in (7) are explicitly labeled by their respective model orders n and $n - 1$. It is interesting to note that the derivative of the relative entropy of the n-th order model is expressed in terms of an expected loss function evaluated for the $(n-1)$-st order model. This is in contrast to (5) and (6), in which all distributions and expectations are with respect to the same model order n (hence the superscript n is omitted in the formulas).

Propositions 2 and 3 both have their advantages. Proposition 2 expresses the derivative of the relative entropy as the correlation between the output and a certain loss function (also a Bregman divergence) in terms of the *same* binomial model. Proposition 3 has to use binomial models of different orders, but directly equates the derivative of the relative entropy to the average of a Bregman divergence.

The remainder of this section proves Propositions 1 and 2.

Lemma 1. *Let P_Y be the pmf of the output of the binomial model described by* (2)*, where the input follows distribution P_X with zero mass on $(-\infty, a]$. For every $y = 0, \dots, n$,*

$$\frac{\mathrm{d}}{\mathrm{d}a} P_Y(y) = \frac{1}{a}\left(y P_Y(y) - (y+1) P_Y(y+1)\right) \tag{9}$$

where we use the convention that $P_Y(n+1) = 0$. The result remains true if P_Y is replaced by Q_Y in (9)*.*

Proof: We start with

$$P_Y(y) = \mathsf{E}\left\{\binom{n}{y}\left(\frac{a}{X}\right)^y \left(1 - \frac{a}{X}\right)^{n-y}\right\}. \tag{10}$$

Evidently,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}a} P_Y(y) \\ &= \mathsf{E}\left\{\binom{n}{y} \frac{\mathrm{d}}{\mathrm{d}a}\left(\left(\frac{a}{X}\right)^y \left(1 - \frac{a}{X}\right)^{n-y}\right)\right\} && (11) \\ &= \mathsf{E}\left\{\binom{n}{y}\left(\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{a}{X}\right)^y\right)\left(1 - \frac{a}{X}\right)^{n-y}\right\} \\ &\quad + \mathsf{E}\left\{\binom{n}{y}\left(\frac{a}{X}\right)^y \frac{\mathrm{d}}{\mathrm{d}a}\left(1 - \frac{a}{X}\right)^{n-y}\right\} && (12) \\ &= \frac{y}{a} \mathsf{E}\left\{\binom{n}{y}\left(\frac{a}{X}\right)^y \left(1 - \frac{a}{X}\right)^{n-y}\right\} \\ &\quad - \frac{n-y}{a} \mathsf{E}\left\{\binom{n}{y}\left(\frac{a}{X}\right)^{y+1} \left(1 - \frac{a}{X}\right)^{n-y-1}\right\} && (13) \\ &= \frac{y}{a} P_Y(y) \\ &\quad - \frac{y+1}{a} \mathsf{E}\left\{\binom{n}{y+1}\left(\frac{a}{X}\right)^{y+1} \left(1 - \frac{a}{X}\right)^{n-y-1}\right\}. && (14) \end{aligned}$$

We note that (11)–(13) hold for $y = 0, \dots, n$. In arriving at (14), we use (10) and the convention that $\binom{n}{n+1} = 0$. In fact the second term in (13) and the second term in (14) are both equal to 0 for $y = n$. Using (10) again, we arrive at (9) from (14).

Since (9) holds for any input distribution P_X, it remains true if P_X is replaced by another distribution Q_X, as long as the input is always greater than a. ■

Lemma 1 is in reminiscent of a result for Gaussian models in [1], where the derivative with respect to the scaling parameter translates to the derivative with respect to the output variable. For the binomial model, the output is discrete and the result consists of the difference of the output distribution (modulated by the variable y) in lieu of derivative.

We next prove Proposition 2.

Proof of Proposition 2: From

$$\mathsf{D}\left(P_Y \| Q_Y\right) = \sum_{y=0}^{n} P_Y(y) \log \frac{P_Y(y)}{Q_Y(y)}, \tag{15}$$

[1]The model used in this paper is equivalent to that in [6], [9]. Specifically, the input to the model in [6], [9] is the inverse of the input X in this paper. Proposition 3 is presented in terms of the model in this paper.

it is not difficult to show that

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}a}\mathsf{D}\left(P_Y\|Q_Y\right) \\
&= \sum_{y=0}^{n}\left(\log\frac{P_Y(y)}{Q_Y(y)}\right)\frac{\mathrm{d}P_Y(y)}{\mathrm{d}a} - \frac{P_Y(y)}{Q_Y(y)}\frac{\mathrm{d}Q_Y(y)}{\mathrm{d}a} \qquad (16)\\
&= a^{-1}(A-B) \qquad (17)
\end{aligned}
$$

where

$$
\begin{aligned}
A &= a\sum_{y=0}^{n}\left(\log\frac{P_Y(y)}{Q_Y(y)}\right)\frac{\mathrm{d}P_Y(y)}{\mathrm{d}a} \qquad (18)\\
&= \sum_{y=0}^{n}\left(\log\frac{P_Y(y)}{Q_Y(y)}\right)(yP_Y(y)-(y+1)P_Y(y+1)) \qquad (19)\\
&= \sum_{y=1}^{n}\left(\log\frac{P_Y(y)}{Q_Y(y)}\right)yP_Y(y) \\
&\quad - \sum_{y=0}^{n-1}\left(\log\frac{P_Y(y)}{Q_Y(y)}\right)(y+1)P_Y(y+1)) \qquad (20)\\
&= \sum_{y=1}^{n}\left(\log\frac{P_Y(y)}{Q_Y(y)}\right)yP_Y(y) \\
&\quad - \sum_{y=1}^{n}\left(\log\frac{P_Y(y-1)}{Q_Y(y-1)}\right)yP_Y(y) \qquad (21)\\
&= \sum_{y=1}^{n} yP_Y(y)\log\frac{P_Y(y)Q_Y(y-1)}{P_Y(y-1)Q_Y(y)} \qquad (22)
\end{aligned}
$$

and

$$
\begin{aligned}
B &= a\sum_{y=0}^{n}\frac{P_Y(y)}{Q_Y(y)}\frac{\mathrm{d}Q_Y(y)}{\mathrm{d}a} \qquad (23)\\
&= \sum_{y=0}^{n}\frac{P_Y(y)}{Q_Y(y)}(yQ_Y(y)-(y+1)Q_Y(y+1)) \qquad (24)\\
&= \sum_{y=1}^{n} yP_Y(y) - \sum_{y=0}^{n-1}\frac{P_Y(y)}{Q_Y(y)}(y+1)Q_Y(y+1)) \qquad (25)\\
&= \sum_{y=1}^{n} yP_Y(y) - \sum_{y=1}^{n}\frac{P_Y(y-1)}{Q_Y(y-1)}yQ_Y(y) \qquad (26)\\
&= \sum_{y=1}^{n} yP_Y(y)\left(1-\frac{P_Y(y-1)Q_Y(y)}{P_Y(y)Q_Y(y-1)}\right). \qquad (27)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&P_Y(y-1) \\
&= \mathsf{E}\left\{\binom{n}{y-1}\left(\frac{a}{X}\right)^{y-1}\left(1-\frac{a}{X}\right)^{n-y+1}\right\} \qquad (28)\\
&= \mathsf{E}\left\{\frac{y}{n-y+1}\left(\frac{a}{X}\right)^{-1}\left(1-\frac{a}{X}\right)P_{Y|X}(y|X)\right\} \qquad (29)\\
&= \frac{y}{n-y+1}\mathsf{E}\left\{\frac{X}{a}-1\,\middle|\, Y=y\right\}P_Y(y). \qquad (30)
\end{aligned}
$$

Taking similar steps leads to,

$$
\frac{Q_Y(y-1)}{Q_Y(y)} = \frac{y}{n-y+1}\mathsf{E}_Q\left\{\frac{X}{a}-1\,\middle|\, Y=y\right\}. \qquad (31)
$$

Therefore,

$$
\frac{P_Y(y-1)Q_Y(y)}{P_Y(y)Q_Y(y-1)} = \frac{\mathsf{E}\left\{X|\,Y=y\right\}-a}{\mathsf{E}_Q\left\{X\mid Y=y\right\}-a}. \qquad (32)
$$

Plugging (32) into (22) and (27) and subsequently (17), we have

$$
\frac{\mathrm{d}}{\mathrm{d}a}\mathsf{D}\left(P_Y\|Q_Y\right) = a^{-1}\sum_{y=1}^{n} yP_Y(y)(T(y)-1-\log T(y)) \qquad (33)
$$

where $T(y)$ is a shorthand for the function defined as the right hand side of (32). By definition (3), we have established (6) in Proposition 2. ■

Proof of Proposition 1: First, fix $x\in(a,\infty)$ and let $Q_Y = P_{Y|X=x}$, which can be regarded as the output distribution of the binomial model with deterministic input a/x, or with the input distribution Q_X being a point mass at x. Evidently, $\mathsf{E}_Q\left\{X\mid Y\right\}\equiv x$. Using Proposition 2,

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}a}\mathsf{D}\left(P_{Y|X=x}\|P_Y\right) \\
&\quad= \frac{\mathrm{d}}{\mathrm{d}a}\mathsf{D}\left(Q_Y\|P_Y\right) \qquad (34)\\
&\quad= \mathsf{E}\left\{\frac{Y}{a}\cdot g\left(\frac{x-a}{\mathsf{E}\left\{X|\,Y\right\}-a}\right)\middle|\, X=x\right\}. \qquad (35)
\end{aligned}
$$

Since (35) holds for every $x\in(a,\infty)$ and

$$
I(X;Y) = \int \mathsf{D}\left(P_{Y|X=x}\|P_Y\right)\mathrm{d}P_X(x), \qquad (36)
$$

averaging both sides of (35) over x according to P_X yields (5). ■

III. The Negative Binomial Model

The negative binomial distribution is defined by the following pmf:

$$
\mathsf{P}\left(Y=y\right) = \binom{y+r-1}{y}(1-q)^r q^y, \quad y=0,1,\dots \qquad (37)
$$

which is the probability that y successful trials are seen before the r-th failure is observed, where the trials are independent Bernoulli trials each with probability q to succeed. We denote this distribution as NB(r,q).

Let b be any positive parameter. With some hindsight, we define a negative binomial model based on random transformation from nonnegative random variable X to random variable Y, where, conditioned on $X=x$, Y has distribution NB$(r, b/(b+x))$. That is, the random transformation is given by conditional pmf

$$
p^r_{Y|X}(y|x) = \binom{y+r-1}{y}\left(\frac{x}{b+x}\right)^r\left(\frac{b}{b+x}\right)^y \qquad (38)
$$

where $y=0,1,\dots$.

The probability distributions, expectations and conditional expectations are similarly defined as in Section II, which depends on the order of the negative binomial model (r) and shall be labeled with superscript r whenever necessary.

Proposition 4. *Let X be an arbitrary nonnegative random variable. Conditioned on X, let Y be a negative binomial random variable with parameters $(r, b/(b+X))$. Then for every $b > 0$,*

$$\frac{\mathrm{d}}{\mathrm{d}b} I(X;Y) = \mathsf{E}\left\{ \frac{Y}{b} \cdot g\left(\frac{X+b}{\mathsf{E}\{X|Y\}+b} \right) \right\}. \tag{39}$$

Proposition 5. *Let P_Y and Q_Y be the output distribution of the binomial model* (38) *induced by input distributions P_X and Q_X, respectively, where P_X and Q_X put no probability mass on $(-\infty, 0)$. Then for every $b > 0$,*

$$\frac{\mathrm{d}}{\mathrm{d}b} \mathsf{D}\left(P_Y \| Q_Y\right) = \mathsf{E}\left\{ \frac{Y}{b} \cdot g\left(\frac{\mathsf{E}\{X|Y\}+b}{\mathsf{E}_Q\{X \mid Y\}+b} \right) \right\}. \tag{40}$$

Again, we compare Propositions 4 and 5 with the following result:

Proposition 6 (Gil Taborda and Pérez-Cruz [6])**.** *Let P, Q, X and Y be defined as in Proposition 5. Then for every $b > 0$,*

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}b} \mathsf{D}\left(P_Y^r \| Q_Y^r\right) \\ &= \frac{r}{b} \mathsf{E}_P^{r+1} \left\{ -\ell(-\mathsf{E}_P^{r+1}\{bX|Y\}, -\mathsf{E}_Q^{r+1}\{bX|Y\}) \right\}. \end{aligned} \tag{41}$$

The distributions and expectations in (41) are explicitly labeled by their respective model orders r and $r+1$.

The contrast between (40) and (41) is similar to the findings in Section II. All quantities in (40) are evaluated with respect to the same negative binomial model; whereas the model orders of the two sides of (41) differ by 1. Proposition 6 equates the derivative of the relative entropy to the average of a Bregman divergence (like $\ell(x, \hat{x})$, the function $-\ell(-x, -\hat{x})$ is also a Bregman divergence).

The remainder of this section is devoted to proving Propositions 4 and 5.

Lemma 2. *Let P_Y be the pmf of the output of the negative binomial model described by* (38)*, where the input is always positive and follows distribution P_X. For every $y = 0, 1, \dots$,*

$$\frac{\mathrm{d}}{\mathrm{d}b} P_Y(y) = \frac{1}{b}\left(yP_Y(y) - (y+1)P_Y(y+1) \right). \tag{42}$$

The result remains true if P_Y is replaced by Q_Y in (42).

Proof: We start with

$$P_Y(y) = \mathsf{E}\left\{ \binom{y+r-1}{y} \left(\frac{X}{b+X}\right)^r \left(\frac{b}{b+X}\right)^y \right\}. \tag{43}$$

Evidently,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}b} P_Y(y) \\ &= \mathsf{E}\left\{ \binom{y+r-1}{y} \frac{\mathrm{d}}{\mathrm{d}b} \left(\left(\frac{X}{b+X}\right)^r \left(\frac{b}{b+X}\right)^y \right) \right\} &(44) \\ &= \mathsf{E}\Bigg\{ \binom{y+r-1}{y} \Bigg[\left(\frac{\mathrm{d}}{\mathrm{d}b} \left(\frac{X}{b+X}\right)^r \right) \left(\frac{b}{b+X}\right)^y \\ &\qquad + \left(\frac{X}{b+X}\right)^r \frac{\mathrm{d}}{\mathrm{d}b} \left(\frac{b}{b+X}\right)^y \Bigg] \Bigg\} &(45) \\ &= \mathsf{E}\Bigg\{ \binom{y+r-1}{y} \left(\frac{X}{b+X}\right)^r \left(\frac{b}{b+X}\right)^y \\ &\qquad \times \left(\frac{-r}{b+X} + \frac{yX}{b(b+X)} \right) \Bigg\} &(46) \\ &= \mathsf{E}\left\{ \binom{y+r-1}{y} \left(\frac{X}{b+X}\right)^r \left(\frac{b}{b+X}\right)^y \left(\frac{y}{b} - \frac{y+r}{b+X} \right) \right\} &(47) \\ &= \frac{y}{b} P_Y(y) - \frac{y+1}{b} \mathsf{E}\left\{ \binom{y+r}{y+1} \left(\frac{X}{b+X}\right)^r \left(\frac{b}{b+X}\right)^{y+1} \right\} &(48) \\ &= \frac{1}{b}\left(yP_Y(y) - (y+1)P_Y(y+1) \right). &(49) \end{aligned}$$

Since (42) holds for any input distribution P_X, it remains true if P_X is replaced by another distribution Q_X, as long as the input is always nonnegative. ■

It is interesting to see that (42) is identical to (9).

The proof of Proposition 5 based on Lemma 2 resembles that of Proposition 2.

Proof of Proposition 5: Using similar techniques as in the proof of Proposition 5, we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}b} \mathsf{D}\left(P_Y \| Q_Y\right) = \frac{1}{b} \sum_{y=1}^{\infty} y P_Y(y)(T(y) - 1 - \log T(y)) \tag{50}$$

where

$$T(y) = \frac{P_Y(y-1)Q_Y(y)}{P_Y(y)Q_Y(y-1)}. \tag{51}$$

Moreover,

$$\begin{aligned} &P_Y(y-1) \\ &\quad = \mathsf{E}\left\{ \binom{y+r-2}{y-1} \left(\frac{X}{b+X}\right)^r \left(\frac{b}{b+X}\right)^{y-1} \right\} &(52) \\ &\quad = \mathsf{E}\left\{ \frac{y}{y+r-1} \cdot \frac{b+X}{b} P_{Y|X}(y|X) \right\} &(53) \\ &\quad = \frac{y}{y+r-1} \mathsf{E}\left\{ 1 + \frac{X}{b} \,\middle|\, Y = y \right\} P_Y(y). &(54) \end{aligned}$$

Similarly,

$$\frac{Q_Y(y-1)}{Q_Y(y)} = \frac{y}{y+r-1} \mathsf{E}_Q\left\{ 1 + \frac{X}{b} \,\middle|\, Y = y \right\}. \tag{55}$$

Therefore,

$$T(y) = \frac{\mathsf{E}\left\{X|\,Y=y\right\}+b}{\mathsf{E}_Q\left\{X\mid Y=y\right\}+b}. \tag{56}$$

Proposition 5 is thus established using (3), (50) and (56). ■

Proof of Proposition 4: The proof is similar to that of Proposition 1. First, fix $x>0$ and let $Q_Y = P_{Y|X=x}$. Using Proposition 5,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}b}\mathsf{D}\left(P_{Y|X=x}\|P_Y\right)\\ &\quad= \mathsf{E}\left\{\frac{Y}{b}\cdot g\left(\frac{x+b}{\mathsf{E}\left\{X|\,Y\right\}+b}\right)\middle|\, X=x\right\}.\end{aligned} \tag{57}$$

Since (57) holds for every $x>0$, averaging both sides of (57) over x according to P_X yields (39). ■

IV. Concluding Remarks

The new information–estimation relationships over binomial and negative binomial models presented in this paper complement the results in [6]. Since the acceptance of this paper, a more comprehensive treatment of information–estimation relationships over binomial and negative binomial models has been developed in [9].

A somewhat related but different random transformation, called *thinning*, has been studied in, e.g., [10]. Whether similar information–estimation relationships as found here can be developed for the thinning models is an open question.

References

[1] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[2] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 1837–1849, May 2008.

[3] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Trans. Inform. Theory*, vol. 58, pp. 1302–1318, 2012.

[4] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inform. Theory*, vol. 56, pp. 3712–3720, Aug. 2010.

[5] T. Weissman, "The relationship between causal and noncausal mismatched estimation in continuous-time AWGN channels," *IEEE Trans. Inform. Theory*, vol. 56, pp. 4256–4273, Sept. 2010.

[6] C. Gil Taborda and F. Perez-Cruz, "Mutual information and relative entropy over the binomial and negative binomial channels," in *Proc. IEEE Int. Symp. Inform. Theory*, pp. 696–700, July 2012.

[7] D. Guo, S. Shamai, and S. Verdú, *The Interplay between Information and Estimation Measures*. Foundations and Trends in Signal Processing, NOW Publishers, 2013. submitted.

[8] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR computational mathematics and mathematical physics*, vol. 7, no. 3, pp. 200–217, 1967.

[9] C. Gil Taborda, F. Pérez-Cruz, and D. Guo, "Information–estimation relationships over the binomial and negative binomial models," *submitted to IEEE Trans. Inform. Theory*, 2013.

[10] O. Johnson and Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality," vol. 56, no. 11, pp. 5387–5395, 2010.