Supplements to

"Optimal Keyword Bids in Search-Based Advertising with Stochastic Advertisement Positions"

Journal of Optimization Theory and Applications

Susan Cholette, Özgür Özlük and Mahmut Parlar

Appendix A Beta Distribution

The beta distribution is used extensively from biological and medical testing, such as tumor detection (Paul [27, p. 423]) to estimating task durations in project management (Keefer and Verdini [28]). We choose the beta density for the ad position because it is one of the most flexible densities for modeling an event constrained by two endpoints, i.e., 0 as the top of the page and 1 as the bottom of the page. A continuous density will be an approximation as the number of sponsored ads on a page is a discrete number, which varies, but typically ranges between 8 to 12 such slots.

In our case the decision variable b and the parameter a are both assumed continuous. See Figure 7 for a 3-dimensional plot of the beta density when $a = 20$. Here, for small (large) values of b, the density is left- (right)-skewed meaning that as b increases the probability of finding the ad near the top of the page also increases. Figure 8 is another 3-dimensional plot of the beta when we fix $b = 10$ cents/bid. In this case, small (large) values of a, the density is right- (left)-skewed. As these figures illustrate, beta density allows substantial flexibility in the desired shape of the ad position distribution X by a suitable choice of the parameter a and the decision variable bid price b.

While there may be thousands of firms bidding on the same keyword, it is impossible to know a priori how much will be bid by each firm. For this reason, given the bid price b, we assumed the ad position to be a beta random variable $X(b)$. But different levels of competition for the same keyword may result in different shapes of the beta distribution. We use the parameter a to be a proxy for the competitive landscape in the bidding process with high values of a corresponding to high levels of competition. For large values of a, the beta distribution would shift to the right. For example, when competition is high, i.e., when other firms have higher valuations than us and thus bid high, we are more likely to see our ad near the bottom of the page as in Figure 8 when a is near 20. Similarly, when competition is low, it is more probable for our ad to appear near the top of the page, again, as in Figure 8 but when a gets closer to 0.

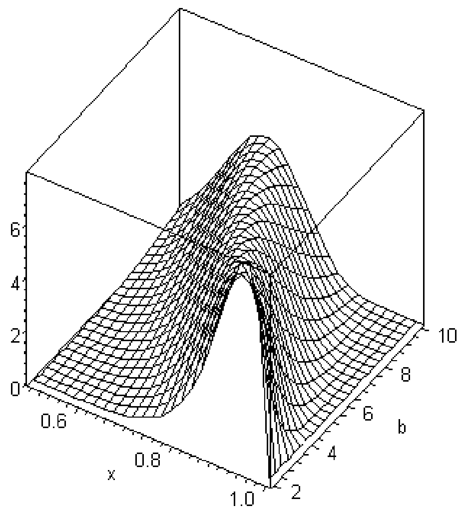

Figure 7: Three dimensional graph of the beta density $f_X(x;b)$ when $a = 20$. Note that for small (large) values of b, the density is left- (right)-skewed.

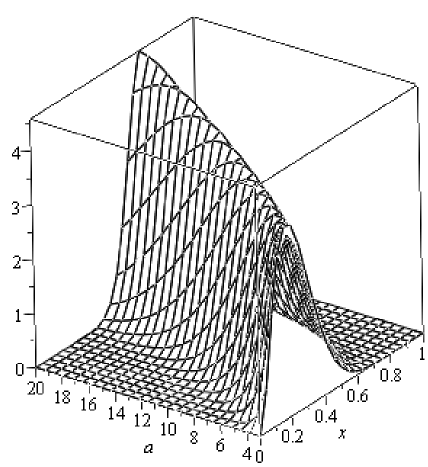

Figure 8: Three dimensional graph of the beta density $f_X(x;b)$ when $b = 10$. Note that for small (large) values of a, the density is right- (left)-skewed.

The beta random variable X has mean and variance,

$$\mu_X \equiv E(X) = \frac{a}{a+b} \quad \text{and} \quad \sigma_X^2 = \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}, \tag{8}$$

and coefficient of skewness $\gamma_X = -[2(a-b)\sqrt{a+b+1}]/[\sqrt{ab}(a+b+2)]$. Note that, if $\gamma_X < 0$, beta density is skewed to the left, so there is a lower probability of ending up near the top. In our model the lower values of x (near 0) places us closer to the top of the page.

Appendix B Factors Affecting the Expected Revenue $\mathcal{R}(b)$

B.1 Number of Impressions per Time $IPT = k$ — A constant

This quantity (i.e., how often the ad is displayed per day) which has dimensions [impr/time], is a function of the popularity of the keyword since certain keywords are more likely to be displayed than others. For example, suppose an advertiser bids on the keyword "Rockport." When a visitor to the search engine website types that keyword into the search box, the advertiser's ad is displayed in the search results, resulting in an *impression*. Since IPT is not a function of bid price decision b, nor is it a function of the ad position X, we assume that given the time series for the impression of a keyword, we use a (constant) point estimate k for the IPT for the time period (i.e., the day) of concern. Most search engines provide a safety mechanism to prevent further impressions of that ad, should the daily budget be vastly exceeded.

B.2 "Click-thru-Rate" CTR (Number of Clicks per Impression) — A Random Variable

The number of clicks per impression (also known as the "click-thru-rate"), CTR, has dimensions [click/impr]. Because the ad position is determined by the bid price, it is reasonable to assume that this r.v. depends on the ad position X, i.e., closer is the ad position to the top of the page (near $x = 0$) the more clicks our ad receives. Thus, we assume that the conditional r.v. $Y(x) \equiv (CTR \mid X = x)$ is a Bernoulli r.v. with parameter $p(x)$, i.e., it has density $f_Y(y \mid x) = [p(x)]^y[1-p(x)]^{y-1}$ with mean $\mu_Y(x) = p(x)$ and variance $\sigma_Y^2(x) = p(x)[1-p(x)]$, both functions of the ad position. In this context it would be convenient to interpret $p(x)$ as the probability of getting a click when the ad appears at position x.

As a reasonable functional form for the probability parameter of the conditional CTR r.v., we assume that $p(x) = (1-x)^m$ for $m \geq 1$, with $p'(x) = -m(1-x)^{m-1} < 0$, $p(0) = 1$ and $p(1) = 0$. This satisfies the requirement that closer the ad is to the top of the page, the higher is the probability that it will be clicked. Later, we will relate the $Y(x)$ r.v. to the number of clicks per time r.v. CPT which will result in a binomial r.v.

B.3 Number of Clicks per Time CPT — A Random Variable

Considered in isolation, the constant impressions per time (IPT) and the r.v. click-thru-rate (CTR) are not very helpful in computing the expected revenue $R(b)$. However, if we write $Y_j(x) \equiv (CTR \mid X = x)$ as the conditional CTR for the jth impression given the ad position, and combine IPT and $Y_j(x)$ as $V = \sum_{j=1}^{IPT} Y_j(x)$, we obtain the "number of clicks per time" (CPT) ([click/time]) as a binomial r.v. with parameters $(IPT, p(x))$, and mean $IPT \cdot p(x)$. Now, since the binomial is accurately approximated by the normal provided that $IPT \cdot p(x) \geq 10$ and $IPT \cdot [1 - p(x)] \geq 10$, we make use of this result and assume that the conditional r.v. $U \equiv (V \mid X = x) = \left(\sum_{j=1}^{IPT} Y_j(x) \mid X = x\right)$ is normal with mean $\mu_U(x) = IPT \cdot p(x)$ and variance $\sigma_U^2(x) = IPT \cdot p(x) \cdot q(x)$ where $q(x) = 1 - p(x)$. (We implicitly assume that the parameter IPT is reasonably large so that the two conditions given above are satisfied.) Since $p(0) = 1$, this implies that when the ad position is near the top ($x = 0$), we expect each impression to result in a click on our ad. To simplify the notation in subsequent discussion, we denote $k \equiv IPT$.

We calculate the (unconditional) expected number of clicks per time, $E(V)$ as,

$$\begin{aligned} E(V) &= \int_0^1 E(V \mid X = x) f_X(x; b)\ dx = \int_0^1 k(1-x)^m f_X(x; b)\ dx = k\frac{\Gamma(a+b)\Gamma(b+m)}{\Gamma(b)\Gamma(a+b+m)} \\ &= kG(b), \end{aligned}$$

where the last line follows from (1).

In a recent paper, Richardson, Dominowska and Ragno [24] indicate that, in general, realized ad positions that are higher (i.e., closer to 1 and thus less desirable) are associated with higher variability CTR's, and thus higher variability CPT's. We represent the variability of CPT conditional on the ad position in terms of its coefficient of variation (c.o.v.) $cv(x) = \sigma_U(x)/\mu_U(x)$ where $\mu_U(x) = kp(x)$ and $\sigma_U^2(x) = kp(x)q(x)$. Thus, the c.o.v. can be written as,

$$cv(x) = \frac{\sqrt{kp(x)q(x)}}{kp(x)} = \sqrt{\frac{q(x)}{kp(x)}}, \tag{9}$$

with $cv(0) = 0$ and $cv(1) = \infty$. Differentiating $cv(x)$, we find that $cv(x)$ is an increasing function of the ad position since,

$$cv'(x) = -\frac{1}{2}\frac{-q'(x)p(x) + q(x)p'(x)}{\sqrt{\frac{q(x)}{kp(x)}}k[p(x)]^2} > 0.$$

The result follows since with $p(x) = (1-x)^m$ and $q(x) = 1 - p(x)$, we have $p'(x) < 0$ and $q'(x) > 0$. This shows that, as suggested by Richardson, Dominowska and Ragno [24], we also find the c.o.v. of CPT to be an increasing function of the ad position.

Rather than working with the complicated $cv(x)$ function in (9) which approaches infinity as x approaches 1, we choose a simpler approximation and assume that $cv(x) = \ell x^n$ where $\ell > 1$ and $n \geq 1$ and where $cv(0) = 0$ and $cv(1) = \ell$. This implies that the standard deviation of the CPT conditional on the ad position is $\sigma_U(x) = k\ell(1-x)^m x^n$.

B.4 Revenue per Click RPC — A Random Variable

The revenue-per-click, RPC, has dimensions [¢/click] and it is also a r.v. denoted by W; we assume that it does not depend on the ad position X. Thus, in this case $W \equiv RPC$ is assumed to have density $f_W(w)$ with constant mean μ_W and variance σ_W^2. In the computation of the expected revenue $R(b)$ we will only need the mean value μ_W, thus we do not make any specific assumptions about the form of the density $f_W(w)$, nor the variance σ_W^2.

To calculate the expected profit we need to determine the expected revenue and expected cost as a function of the bid price b. In our problem, the revenue per time is the product of (i) the number of clicks per time, and (ii) the revenue per click, i.e.,

$$R \equiv \left(\sum_{j=1}^{IPT} Y_j(x)\right) \cdot RPC = V \cdot W, \tag{10}$$

with dimensions [¢/time].