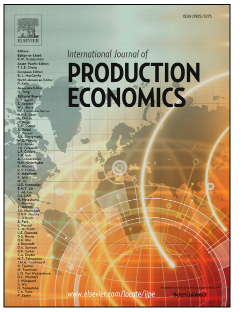
Accepted Manuscript



A generalized count model on customers' purchases in O2O market

Ruixia Shi, Hongyu Chen, Suresh P. Sethi

PII: S0925-5273(17)30355-9

DOI: 10.1016/j.ijpe.2017.11.009

Reference: PROECO 6872

To appear in: *International Journal of Production Economics*

Received Date: 1 January 2017

Revised Date: 25 June 2017

Accepted Date: 2 November 2017

Please cite this article as: Shi, R., Chen, H., Sethi, S.P., A generalized count model on customers' purchases in O2O market, *International Journal of Production Economics* (2017), doi: 10.1016/j.ijpe.2017.11.009.

A Generalized Count Model on Customers' Purchases in O2O Market

Ruixia Shi
School of Business, University of San Diego, San Diego, CA 92110 rshi@sandiego.edu

Hongyu Chen
College of Business Administration, California State University, Long Beach, Long Beach, CA 90840 Hongyu.chen@csulb.edu

Suresh P. Sethi
Jindal School of Management, SM 30, University of Texas at Dallas, P.O. Box 830688, Richardson, TX, 75083
sethi@utdallas.edu

A Generalized Count Model on Customers' Purchases in O2O Market

ABSTRACT

In the Online-to-Offline (O2O) ecommerce model, one challenge facing the online business is to predict customers' future purchases towards each product or subcategory of products, and consequently, coordinate the large amount of offline businesses involved. The main obstacle in doing that originates from the highly diversified services and thus the customer base which offline businesses bring in. The heterogeneity of customers, geographic or demographic, needs to be accurately accounted for. However, although the previous transactions for each customer are well documented, his/her demographic data is difficult or costly to acquire. Traditional wisdom relies on fitting customers into some specific statistical distribution to arrive at a satisfactory stochastic model, which may be accurate, to some extent, at a higher level. This is the case for the classic Beta-Binomial/Negative Binomial Distribution (BB/NBD) model on customers' repeat purchasing in offline context. Nevertheless, to deal with the complex level in customers' heterogeneity at an O2O business, using specific distribution is inadequate, let alone the mathematical challenges.

We propose a new model to deal with the diversity of customers. Using BB/NBD as a starting point, we relax the Beta assumption in the model to include a generalized distribution. The generalization is made possible through using the Gaussian quadrature. The results retain the elegance of stochastic model while at the same time it captures customers' heterogeneity at a better, granular level. We use a dataset from Ctrip.com, a leading O2O provider in China, to show that the proposed method outperforms the BB/NBD model in both in-sample and out-of-sample predictive performance. Our approach provides a practical solution for O2O practitioners to forecast their future demands.

Keywords: Business Analytics, Online-to-Offline, Repeat Purchase, BB/NBD, Count Model

1. Introduction

In a typical Online-to-Offline (O2O) business, customers purchase some service online (e.g., order a meal, book a hotel room or a trip through APPs using their cell phones), and consume the service offline through a third party. O2O has several differences with the traditional Business-to-Customer (B2C) model. First, in O2O model, the online business coordinates with the third party to provide service offline. A B2C business, however, provides products to its customers directly. On the other hand, an O2O business is more than an agent or match maker. The online business and the offline service providers take separate roles in the value chain, with online business providing website store front, advertising, order processing and customer support (like hosting online reviews, resolving customers' issues etc.). The service provider focuses on its core business. In this way, O2O model lowers the bar for offline businesses to enter the online market. Second, the scope of products provided in O2O market is different or wider than the B2C. A B2C business usually deals with physical products. In O2O model, however, offline providers could sell not only products but also their services (e.g., a meal in a restaurant or a stay at a hotel). After smartphones penetrated the world, service products are now "just one touch away", which leads to the tremendous growth of O2O market.

Challenges are accompanied by benefits. An O2O online business needs to coordinate tens of thousands offline businesses. It needs to effectively monitor these providers, allocate the capacity accordingly and deliver the desired customer service. Among them, capacity allocation/planning is probably the most important/challenging task for the online business. Take Ctrip.com (one of the leading O2O websites in China) as an example. If Ctrip can predict the future total purchases on a specific hotel (e.g., Sheraton) for the next several months, or more generally, customers' future purchases on a specific sub-type of service (e.g., customers' spending on meals), it will greatly facilitate the communication with the providers (in this case, Sheraton hotel) or promotion strategies (meal products vs. hotel products). In this paper, we aim to provide a practical solution on: *How to predict customers' future purchases in the O2O market?*

One advantage of the O2O market is that online businesses have documented all the transactions of customers. Since many products/services purchases tend to be repetitive (e.g., haircut, meal, etc.) in O2O, using historical data to predict future actions is a quintessential business analytics (BA) problem.

Online businesses usually do not know customers' demographic information (or it is very difficult/costly to acquire), except customers' purchase history. Regression based methods thus will be difficult to implement without a further understanding of each customer. For O2O market, customers are highly diverse, because O2O involves a large scope of offline businesses, providing a large variety of services (such as hotel, retreat, restaurant). A typical online business (ctrip.com, meituan.com) can organize thousands of offline businesses in hundreds of categories. These businesses may be geographically located in different places. The diversity of the service and the location leads naturally to the diversity of customers (e.g., the same hotel chain in different cities). Meanwhile, many O2O customers purchase the service in groups, (e.g., a family/group of friends buy a trip, movie tickets. At Ctrip.com, the average number of travelers for each tour order is 2.72). To make the situation worse, even for the same customer account, the travel partners may be different for each trip. These complexities (diversity and grouping) of customers' base need to be effectively addressed to predict customers' future purchases.

In the literature, predicting customers' future purchase in O2O market can be viewed as a special case of the customer repeat purchasing problem, where customers repeatedly shop in a category (e.g., purchasing for tooth paste). A stochastic model --the Beta-Binomial/Negative Binomial Distribution (BB/NBD) model is probably the most well-known framework used in this situation (Jeuland et al. 1980, Morrison and Schmittlein 1988, Winkelmann 2008). BB/NBD model assumes that a customer's categorical spending follows a Negative Binomial Distribution (NBD), while purchases of a focal brand are Binomially distributed. To capture customers' heterogeneity across population, the selection rate of the focal brand at each purchase follows a Beta distribution. Many previous studies have confirmed the validity of the BB/NBD model (Morrison and Schmittlein 1988, Fader and Hardie 2000).

One concern for BB/NBD model is that it oversimplifies the reality, especially the assumption of Beta distribution. In many cases, Beta distribution is flexible enough to capture customers' variations, and thus is well accepted in modeling customers' heterogeneity (Chatfield and Goodhardt 1970, Morrison and Schmittlein 1988, Fader and Hardie 2010). But in the case of O2O business, since customers' base is highly diverse and may result in more than one mode in heterogeneity distributions or even spikes in the distributions, Beta distribution will fall short to fully capture the heterogeneity among customers.

In this paper, we propose a new approach to capture customers' heterogeneity using a general distribution. Specifically, we assume customers' selection rates may follow any arbitrary distribution, which is a big step forward on the original BB/NBD model, and can solve the heterogeneity problem effectively. However, in achieving this, we bring a general function form into our model, and the mathematical representation is no longer tractable. We further introduce the Gaussian quadrature to our model to overcome this difficulty.

Our key contribution is based on the following observation. In the BB/NBD framework, the distribution for customers' selection rates only shows up within the integral. After introducing the Gaussian quadrature to approximate the integral, the untraceable general distribution is reduced to k discretized values. These discretized values could be treated as k unknown "parameters" of the distribution. We could estimate these parameters using maximum likelihood as we estimate the two parameters for Beta distribution in the BB/NBD framework. Since the Gaussian quadrature is accurate for all polynomials up to the degree of $2k - 1$, the whole formula converges exponentially fast as k increases (Press et al. 2007, p180). Thus, we have a customer purchasing model which is general in nature as well as mathematically elegant.

To validate our model, we acquired a dataset from Ctrip.com, one of the leading O2O websites in China. Our results show that the proposed GB/NBD model could effectively address the customer heterogeneity brought by the vast diversifications in the O2O context.

In below, we first review the literature in Section 2. The details of our approach are illustrated in Section 3. We then conduct an empirical analysis and a simulation study to validate our model and evaluate its predictive performance in Sections 4 and 5, respectively. Concluding remarks are presented in Section 6 at the end of the paper.

2. Literature Review

O2O, as a new ecommerce model, has received growing attention in academia in recent years. Researchers have discussed various aspects of O2O, including the reputation management (Xiao and Dong 2015), social relation management (Tsai et al. 2013), supply chain power structure (Chen et al. 2016), service quality (Du and Tang 2014), and advertising/recommendation scheme (Chen et al. 2013). However, to our knowledge, there is a lack of study discussing how to improve the forecasting accuracy specifically in O2O market, or using O2O data.

General forecasting methods have a long research stream. Common forecasting methods include time series analysis, panel data models, machine learning based models, and stochastic models (e.g., BB/NBD model). Time series analysis uses different techniques, such as auto-regression, exponential smoothing, autoregressive integrated moving average (ARIMA) model, etc. These methods are well established, but do not incorporate enough factors, or take individual effect into account. Machine learning based models, one the other hand, demand a larger number of variables at an individual level and have gained its popularity and success in the internet age (Choi et al. 2014). The panel data model and stochastic model lie in between in terms of data needed; they incorporate more factors than time series analysis, but less than machine learning based method. Both Panel data (Ren et al. 2015) and stochastic models (Fader and Hardie 2001, 2005, Abe 2009) have reported success in various business scenarios. We need to point out that each method has its own merit and there is not one method which is superior than others. Complicated models may have higher accuracy, however, as Goldstein and Gigerenzer (2009) pointed out: “simple statistical forecasting rules, which are usually simplifications of classical models, have been shown to make better predictions than more complex rules, especially when the future values of a criterion are highly uncertain.” In O2O case, since customers exhibit a strong repetitive purchase pattern, adapting the well-established NBD based model to this case is both natural and practical.

Using NBD based model to forecast customers' categorical spending was established in 1980s (Morrison and Schmittlein 1988, Schmittlein et al. 1985). Dunn et al. (1983, p. 256) states that "for most purposes in brand purchasing studies, the NBD tends to be accepted as robust to most observed departures from its (stationary) Poisson assumption". Take BB/NBD as an example. Its' main idea is to assume that a customer's categorical spending follows a Negative Binomial Distribution (NBD), while purchases of a focal brand are Binomially distributed. The selection rate of the focal brand at each purchase follows a Beta distribution. The main reason for using Beta distribution to model customers' heterogeneity on selection rates is the mathematical convenience--Beta is the conjugate prior of Binomial distribution. As a result, the BB/NBD model is mathematically tractable and a lot of predictive results can be expressed in closed forms. The whole implementation could be conducted in MS Excel (Morrison and Schmittlein 1988, Fader and Hardie 2005, Zheng et al. 2012). In O2O case, when facing a highly diversified customer base, how could we relax the assumption of Beta distribution within the framework and keep the mathematical simplicity at the same time? To answer this question is the main purpose of the manuscript.

3. The Model

3.1. The Base Model

The canonic story behind our model is that a customer repeatedly visits an O2O website to purchase some services (e.g., purchasing tour service). She may not purchase the same product/service (e.g., a trip to Shanghai, a stay at Hilton chain) on every visit. We discretize her decision process into two steps: to visit the website for a product (e.g., a tour), and to make the purchasing decision, i.e., whether to purchase the focal brand/subcategory (Hilton vs. Sheraton; or a group tour vs. free-travel tour, etc.) or not.

To quantify the first step, we follow Schmittlein et al. (1985) and assume:

i. the number of purchases (n) made by a customer during a time period follows a Poisson distribution with rate λ;

ii. the purchase rate λ across customers follows a gamma distribution, a standard conjugate prior to the Poisson distribution. That is, $f(\lambda) =$

$\alpha^r \lambda^{r-1} e^{-\lambda\alpha} / \Gamma(r)$ with the shape parameter r, the scale parameter α, and $\Gamma(\cdot)$ denoting the gamma function.

Since customers may have different purchase frequencies, to capture the heterogeneity of purchase rates among customers, we introduce the Gamma distribution in assumption *ii*. Assumptions *i* and *ii* jointly prescribe the number of total purchases (N) of a random customer to follow a negative binomial distribution (NBD). That is,

$$P(N = n|r, \alpha) = \frac{\Gamma(r+n)}{\Gamma(r)n!} \left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^n . \qquad (1)$$

Using the NBD distribution to model customers' repeated purchasing has been verified and proved to be very robust for store-level data (Schmittlein et al. 1985, Dunn et al. 1983, Morrison and Schmittlein 1988).

Within the category spending, we assume:

iii. for each visit, the probability that a customer selects the focal brand/ subcategory is *p*. Thus, the number of total purchases of the focal brand/ subcategory by a customer follows a binomial distribution;

iv. selection rate *p* across customers follows a general distribution with pdf $g(p)$;

v. a customer's categorical spending decision and her brand/subcategory choice decision are independent of each other.

The hidden notion behind assumption *iii* is that each customer is stable on her brand preference during the observation period. In reality, customers' favorites at individual level may not be constant over time. Morrison and Schmittlein (1988) refer to this as non-stationarity and provide a detailed discussion in their paper.

Since customers' preference towards the focal brand may be different from each other, they possibly will have different selection rates *p*. To fully capture the heterogeneity among customers, we assume that selection rates follow a general distribution in assumption *iv*. Introducing a general distribution for *p* represents an important departure from the existing studies on customers' repeat purchasing. In the traditional BB/NBD model, the selection rates are assumed to follow a Beta distribution, the conjugate prior of the Binomial distribution. As a result, the whole BB/NBD model is mathematically tractable. The resulting BB/NBD model has proved to be quite robust and has been adopted in solving many business problems (Morrison and Schmittlein 1988, Fader and Hardie 2000, Zheng et al. 2012). However, this assumption is made for mathematical convenience. If there are more than two modes in the selection rates' distribution or there exist spikes, the predictive accuracy of using the BB/NBD model is doubtful. Moreover, it is very difficult to acquire customers' data to test the distributions of their selection rates in practice. For the purpose of generality and to fully capture the heterogeneity among customers, we proceed to assume that p follows a general distribution across the population with pdf $g(p)$ instead of assuming a specific form for p to be the Beta distribution as in the BB/NBD model. Since in our model, the distribution of brand purchasing probability is a general distribution, we term our model as the GB/NBD model throughout the paper.

With assumptions *i* to *v*, we can derive the distribution of a random customer's total number of purchases of the focal brand (X) as:

$$P(X=x)=\iint P(X=x|\lambda,p)f(\lambda)g(p)d\lambda dp=\iint \frac{(\lambda p)^x e^{-\lambda p}}{x!}f(\lambda)g(p)d\lambda dp$$

$$=\int_0^1 \frac{\alpha^r p^x \Gamma(x+r)}{x!\,\Gamma(r)(\alpha+p)^{x+r}}g(p)dp$$

$$=\frac{\alpha^r \Gamma(x+r)}{x!\,\Gamma(r)}\int_0^1 \frac{p^x}{(\alpha+p)^{x+r}}g(p)dp.$$

This distribution involves the integration of $g(p)$. As we know, in numerical analysis, the definite integral of a function, could be stated (through quadrature rule) as a weighted sum of function values at specified points within the domain of integration. If we adopt the Gaussian quadrature formula, the above equation can be expressed as

$$P(X=x)\approx\frac{\alpha^r \Gamma(x+r)}{x!\,\Gamma(r)}\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i), \qquad (2)$$

where k is the order of the Gaussian quadrature, which is chosen based on approximation accuracy. Please note, generating p_i (the position of abscissa), and $w(i)$ (the weight for abscissa i) are a standard process of Gaussian quadrature. The detailed calculation of $w(i)$ is complicate, but fortunately, one can use the standardized numerical procedure to produce them, or directly look them up in a Gaussian quadrature table. $g(p_i)$ is the value of $g(p)$ at the Gaussian abscissa p_i. Gaussian quadrature (with order k) does not have error if the integrand can be expressed in polynomials (with order $2k-1$). For nonpolynomial smooth integrands, the error decreases by a factor of at least 4 with each increase of k to $k+1$. Further discussion on the Gaussian quadrature could be found in Press et al. (2007, p180).

3.2. Parameter Estimation

Recall that in the traditional BB/NBD model, two parameters of the Beta distribution need to be estimated to make predictive results. Under the empirical Bayes framework, the parameters of the BB/NBD model can be estimated using maximum likelihood (MLE). After introducing the Gaussian quadrature in our model, the general distribution $g(p), p \in [0,1]$, is reduced to k values of $g(p_i), i = 1, \ldots, k$. If we treat $g(p_i)$ as k unknowns in Equation (2), we will have a total of $k+2$ unknown parameters ($\alpha, \gamma, g(p_i)$) to be estimated given the customers' purchasing data (X). We also resort to the maximum likelihood for estimating parameters. The detailed procedure of parameter estimation is presented in the following table:

Table 1 Pseudocode for Parameter Estimation

Initialization	Determine the number (J) of customers in the data set Assign a value to k, the order of Gaussian quadrature Set $\alpha = \alpha^0, r = r^0, g(p_i) = g^0(p_i)$, *i* =1,…,*k* Set the log-likelihood value *LL =0*
Iteration	For customer *j* (*j*=1,…,*J*), *1) Count the total number of focal brand purchases* x_j *from the dataset* *2) Calculate the probability that the customer make* x_j *purchases* $P(X{=}x_j)$ *from Equation (2)*

	3) LL= LL + Log[P(X=x_j)]
	4) Move to j+1
	Continue the iteration to cover all customers to get the final LL value
Optimization	*Find the optimal set of $\alpha = \alpha^*, r = r^*, g(p_i) = g^*(p_i)$ which maximizes the final LL value*

For the last optimization step, we use a nonlinear optimization algorithm to obtain the optimal set of the parameters. There is a practical issue. In some situation (when the likelihood function is flat, thus difficult to find the maximum), the searching algorithm may not converge to the correct solution. This is the limitation of all MLE methods. We thus use a convenient approach in which we split the searching into two steps. First, we estimate the BB/NBD model, and feed the parameters of NBD distribution for our GB/NBD model. Second, we estimate the GB/NBD model with k Gaussian parameters (for Beta distribution). Splitting the searching space greatly facilitates the nonlinear solver for MLE.

3.3. Posterior Distributions

After estimating all the parameters, it is easy to predict the customers' purchasing rates, selection rates, and total purchases across all different brands based on assumptions *i* to *v*. We list some of the key results in this section. The proofs and the detailed derivations are all relegated to Appendix.

A brand manager, after observing the total purchases of a focal brand by a customer, can estimate the distribution of total purchases by the customer across different brands as follows:

$$P(N = n|X = x) = \int\int P(N = n|X = x, \lambda, p)h(\lambda, p|X = x)d\lambda dp$$

$$= \frac{\Gamma(n + r)\sum_{i=1}^{k} p_i^x (1 - p_i)^{n-x} g(p_i)w(i)}{(1 + \alpha)^{n+r}(n - x)!\,\Gamma(x + r)\sum_{i=1}^{k} \frac{p_i^x}{(\alpha + p_i)^{x+r}} g(p_i)w(i)}.$$

The expected number of total purchases across different brands conditional on the purchases of the focal brand is expressed as

$$E(N|X=x)=\sum_{n=x}^{\infty} nP(N=n|X=x)$$

$$=\frac{\sum_{i=1}^{k}\frac{x(1+\alpha)+r(1-p_i)}{\alpha+p_i}\left(\frac{p_i}{\alpha+p_i}\right)^{x}\left(\frac{1}{\alpha+p_i}\right)^{r} g(p_i)w(i)}{\sum_{i=1}^{k}\left(\frac{p_i}{\alpha+p_i}\right)^{x}\left(\frac{1}{\alpha+p_i}\right)^{r} g(p_i)w(i)}.$$

We know that

$$\sum_{n=x}^{\infty}\frac{n\Gamma(n+r)(1-p_i)^{n-x}}{\Gamma(x+r)(1+\alpha)^{n+r}(n-x)!}$$
$$=\sum_{n\prime=0}^{\infty}\frac{n'\Gamma(n'+x+r)(1-p_i)^{n'}}{\Gamma(x+r)(1+\alpha)^{n'+x+r}n'!}$$
$$+x\sum_{n\prime=0}^{\infty}\frac{\Gamma(n'+x+r)(1-p_i)^{n'}}{\Gamma(x+r)(1+\alpha)^{n'+x+r}n'!}.$$

By letting $x+r\ =\ r'$, we have

$$E(N|X=x)=\left(\frac{1+\alpha'}{\alpha'(1+\alpha)}\right)^{r'}\left(\frac{r'}{\alpha'}+x\right)=\left(\frac{1}{\alpha+p_i}\right)^{r'}\left(\frac{r'(1-p_i)}{\alpha+p_i}+x\right).$$

After obtaining the model parameters, we also can gain some insights on customers' selection rates of the focal brand, which are given as follows:

$$g(p|X=x)=\int_0^{\infty} h(\lambda,p|X=x)d\lambda=\frac{p^x g(p)}{(\mathrm{p}+\mathrm{\alpha})^{\mathrm{r+x}}\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}.$$

The mean of the selection rate given the total number of purchases of the focal brand can be written as

$$E(p|X=x)=\int_0^1 pg(p|X=x)dp$$
$$=\int_0^1 \frac{p^{x+1}g(p)}{(\mathrm{p}+\alpha)^{\mathrm{r+x}}\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}$$
$$=\frac{\sum_{i=1}^{k}\frac{p_i^{x+1}}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}{\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}.$$

Similarly, we can obtain the marginal distribution of the purchase rate given the total purchases of the focal brand as

$$f(\lambda|X=x)=\int_0^1 h(\lambda,p|X=x)dp=\frac{e^{-\lambda\alpha}\lambda^{r+x-1}\sum_{i=1}^{k}e^{-\lambda p_i}p_i^x g(p_i)w(i)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}.$$

The mean of a customer’s purchase rate given her total number of purchases of the focal brand then is

$$E(\lambda|X=x)=\int_0^\infty \lambda f(\lambda|X=x)d\lambda=\frac{\Gamma(r+x+1)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r+1}}g(p_i)w(i)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}.$$

4. Empirical Illustration

4.1. The Data

We acquired the data from Ctrip.com. Ctrip is the largest online travel agent (OTA) in China with market cap around 20B. It provides a variety of travel services such as group travel, free travel, cruise and ticket service, etc. It, cooperating with Baidu.com, has become a major player of the O2O market in China in recent years. Our data records the purchase transactions of Ctrip on total 1793 travel service products in six categories. The time spans from March 1, 2014 to August 31, 2014, in 6 months. To collect this dataset, we focused on a group of 3458 customers who have been active since January 1, 2014 (made at least one purchase of any product from January 1 – February 28, 2014). The data set is splitted into two parts with each covering three months. Table 2 shows the frequency of customers' purchases on the "free-travel"[1] subcategory from the first three months. The second half (June 2014 – August 2014) are saved for validation of the model's predictive performance. Among the 3458 customers, 1630 of them never purchased any "free-travel" product in the first three months, 410 of customers purchased just 1, 182 customers purchase 2 free-travels, and so on.

Table 2 Ctrip Dataset

Number of purchases	Frequency
0	1630
1	410
2	182
3	80
4	88
…	…
18	37
19	33
>=20	272

4.2. Model Fitting Results

Equation (2) gives the unconditional distribution of a random customer's total purchases of a specific subcategory in a given period of time. As discussed in the section on parameter estimation, we can estimate the parameters and unknowns using the maximum likelihood after observing the customers' purchasing data.

The estimated parameters are presented in Table 3. The parameters for Fader and Hardie (2000) are put under the column BB/NBD. The log-likelihood value (LL) of the estimation is -7278. The estimated parameters for our GB/NBD model are shown in the column GB/NBD. We use a 10^{th} order Gaussian quadrature (k =10) to obtain these results. Recall that a higher LL indicates a higher chance of overall model fitting. From Table 3,

[1] Ctrip offers a variety of products including tours, hotel, air tickets, visa and cruise etc. "Free-travel" is a subcategory of tour service. It only provides lodging and commuting service, which is very similar to group buying.

it is easy to see that our maximum likelihood estimation converges to a better value of -7248 than the BB/NBD.

Table 3 Parameters Estimated

Parameters	BB/NBD	GB/NBD (with k=10)
-LL	-7278	-7248
γ	30.891	3.026
α	1.152	0.208
a	0.185	NA
b	0.815	NA
$g(p_1)$	NA	9.667
$g(p_2)$	NA	0
$g(p_3)$	NA	0
$g(p_4)$	NA	0
$g(p_5)$	NA	0
$g(p_6)$	NA	0
$g(p_7)$	NA	0
$g(p_8)$	NA	0.980
$g(p_9)$	NA	0.916
$g(p_{10})$	NA	15.061

In Table 4, we present the fitted values for both BB/NBD and GB/NBD models. From the table, we can see that both models fit well with the real dataset. The GB/NBD model has smaller χ^2 goodness-of-fit value (13.165 vs. 72.667), which means GB/NBD has a better fit.

Table 4 Comparison of the Fitted Values

Number of purchases	True Frequency	Fitted (BB/NBD)	Fitted (GB/NBD k=10)
0	1630	1651	1618
1	410	307	433
2	182	183	167
3	80	135	106
4	88	108	82
…	…	…	…
18	37	36	36
19	33	34	33

>=20	272	272	267
Goodness-of-fit		72. 669	13.165

Figure 1: Distribution of the Brand Choice Probability (p)

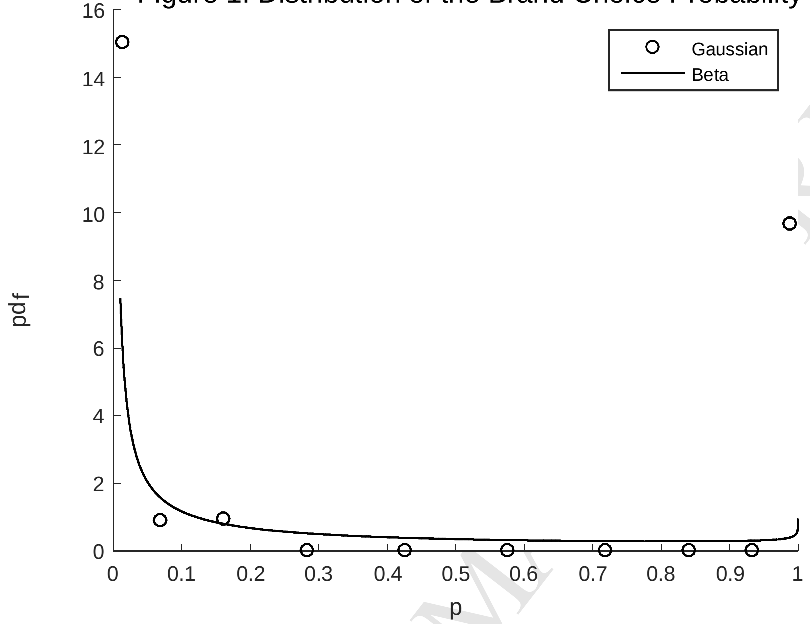


Figure 1 shows the comparison of the estimated customers' brand choice probability by Beta distribution (the solid line) and our proposed general distribution (the circle line). Both curves show that probability density is high at the ends ($p = 1$, and $p = 0$). However, for $0 < \mathrm{p} < 0.3$, the Beta distribution curve is very smooth, while the general distribution curve shows fluctuations.

4.3. Order of Gaussian Quadrature

In the GB/NBD model, the order of Gaussian quadrature (k) needs to be selected before estimating the parameters. As we know, if the underlying distribution $g(p)$ can be approximated by a polynomial function of order *2k-1*, we could use a k^{th} order Gaussian quadrature to accurately approximate $g(p)$ (Stoer and Bulirsch 2002). Thus, a 10^{th} order Gaussian quadrature can approximate polynomial functions of 19^{th} order, which is adequate enough for a majority of common distributions.

Using the Ctrip data (where we do not know the exact form of underline distribution), we test the impacts of different order (k) of Gaussian quadrature on the goodness of fit for GB/NBD model. The results are shown in Table 5. As we can see, for this dataset, any number $k > 6$ gives better fitness results than the BB/NBD model. Computationally, both BB/NBD and the GB/NBD model converge within minutes on a modern computer (Intel i7).

Table 5 Sensitivity on the Order of Gaussian Quadrature

k	LL	Goodness-of-fit
2	-15216	1563.1
3	-7311	134.2
4	-7334	166.3
5	-7308	125.5
6	-7277	75.9
7	-7258	31.9
8	-7254	23.7
9	-7251	17.9
10	-7248	13.2
11	-7247	10.8
12	-7246	9.7

4.4. Predictive Results

Predicting a customer's future-period purchases given her observed current-period activity is often of central interest to any business. We further evaluate our model's predictive performance by computing the conditional expectation of a customer's purchases. To simplify the model estimation, we assume that the second period is of equal length to the first (observed) period. In our case, the Ctrip data is splitted into two equal periods: the first 3 months as the in-sample and the other half as the out-of-sample. We use the first half of Ctrip data to estimate the parameters of the BB/NBD model and

GB/NBD model. After that, using Robbins' results (Robbins 1977), the predicted purchases for the second half could be derived as:

$$E(X'_2|X_1 = x) = (x + 1) * P(X_1 = x + 1)/P(X_1 = x).$$

Where X_1 is the focal customer's purchases in the first half, and X'_2 is the predicted number of purchase in the second half. This equation represents the expected future purchases (on a specific brand of a customer) given her total number of purchases on the focal brand in the first period. The predictions of the two models are then compared with the real sales for each of the 3458 customers in the second half to evaluate the predictive performance.

To measure the difference of the predicted value and the true value, we calculate the *root-mean-square error* (RMSE) of the two series, where a lower RMSE value indicates a better prediction. The *RMSE* values for BB/NBD model is 9.010 and for GB/NBD is 6.267 respectively. This demonstrates superior predictability of our proposed model where the heterogeneity of the customers has been accurately accounted for.

In summary, the model fitness, in-sample model validation and out-of-sample prediction all indicate that the GB/NBD model is effective for real world data. Since the underline mechanism for customers' decision process is unobservable, we next further examine this model through a simulation study, where every step could be verified.

5. Simulation Study

One advantage of the GB/NBD model is that it can accurately approximate any distribution of customer's heterogeneity, which is very difficult to measure in reality. In this section, we conduct a series of simulation study to evaluate our model's performance when the underlying distribution deviates from the Beta distribution. Two representative distributions are tested: truncated normal and piecewise linear. Using these two distributions, we compare the simulation results of our GB/NBD model with the traditional BB/NBD model.

The setup of the simulation for the GB/NBD model is as follows:

1) There are 1,000,000 customers.

2) A purchase rate $\lambda_i \sim \Gamma(\alpha, r), i = 1 \ldots 1{,}000{,}000$ is randomly assigned for each customer.
3) For customer i, $i = 1 \ldots 1{,}000{,}000$, we conduct a random draw from the Poisson distribution with rate λ_i. The resulting number $n_i = \lambda^n e^{-\lambda}/n!$ is assigned as customer i's total category purchases.
4) The selection probability for the focal brand, i.e., $p_i \sim \text{trancated normal}(\mu, \sigma^2)$ or $p_i \sim \text{piecewise linear}, i = 1 \ldots 1{,}000{,}000$ is randomly assigned for each customer.

To get the simulation results for the BB/NBD model, we follow the first three steps for the GB/NBD model and change the distribution of the focal brand selection rate of customer i to a Binomial distribution with parameters n_i and p_i.

We estimate the parameters of the GB/NBD model using Equation (2). For the BB/NBD model, we use Equation (5) from Fader and Hardie (2010):

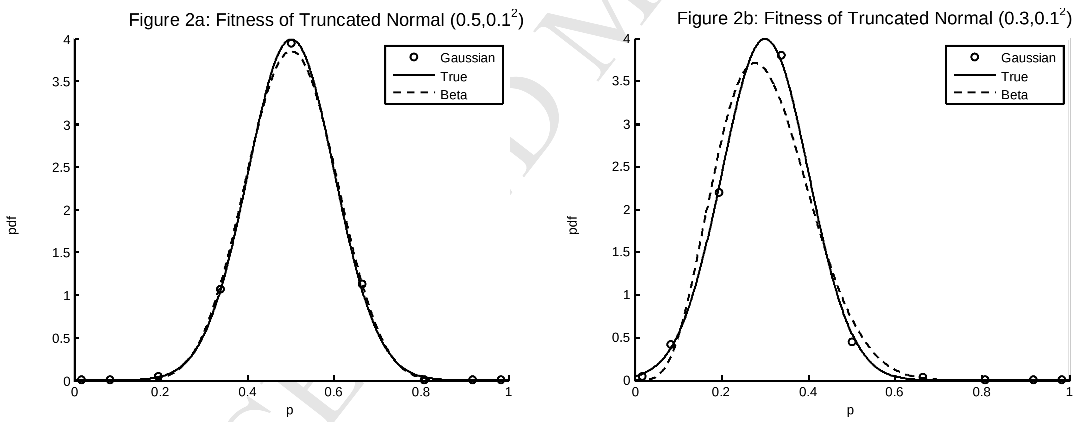
$$P(X = x) = \frac{\Gamma(r + x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x \frac{\Gamma(a + x)}{\Gamma(a)} \frac{\Gamma(a + b)}{\Gamma(a + b + x)} \times {}_2F_1\left(r + x, b; a + b + x; \frac{1}{\alpha + 1}\right)$$

where ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function (Abramowitz and Stegun 1972). A technical issue worth noting is that in the online context, the total number of category purchases n_i for customer i is known by online retailers (e.g., Ctrip.com has each customer's purchasing records of any products including group tours, free travel, tickets, etc.). We could estimate the Gamma parameters (r, α) first based on n_i only. Then, we proceed to estimate the whole model. Since we need to use a nonlinear optimization algorithm to search for the maximal value of the likelihood function, the two-step estimation strategy is less challenging.

In the simulation, we set $r = 1.8$ and $\alpha = 0.09$ for the Gamma distribution. To evaluate the accuracy of predicting the selection rates' distribution, we keep the Gamma distribution unchanged, and vary the parameters of the truncated normal distribution in step 4). Our results are shown in Figures 2a-3b.

In Figure 2a, the solid line is the true distribution (truncated normal distribution with $\mu = 0.5$ and $\sigma = 0.1$); the dashed line is the fitting result of the Beta distribution. As we can see, the Beta line agrees with the true distribution quite well except the part around the mean (from 0.45 to 0.55). The fitted values by GB/NBD are presented by the circle dots in the figure. It is easily seen that the fitted results by GB/NBD agree very well with true distributions. Note that the approximated values of the Gaussian quadrature are shown in scatter dots; this is because only the values at the Gaussian abscissas on the curve matter.

Figure 2b is analogous to Figure 2a. It is initialized with a truncated normal distribution with $\mu = 0.3$ and $\sigma = 0.1$. The accuracy of the fitting for the Beta distribution gets worse. The disagreement of true curve (solid line) and the Beta curve (dashed line) is visually observable. The GB/NBD (circle dots) keeps its performance and agrees with the true curve very well. Note that in Figures 2a and 2b, a 9th order Gaussian quadrature is used.



In Figure 3a, we test an extreme case when the underline distribution is piecewise linear (as shown by the solid curve). The estimated Beta curve fits the true distribution poorly. Compared with the Beta curve, the values predicted by the GB/NBD model are much better. The results in Figure 3a are obtained using a 15th order Gaussian quadrature. We further increase the order of the Gaussian quadrature from 15 to 21, and present the fitting results in Figure 3b[2]. The circle dots obtained from the GB/NBD model clearly track the true curve. These results evidently demonstrate the advantage of the GB/NBD model over the traditional BB/NBD model.

Figure 3a: Approximation of Piecewise Linear Distribution

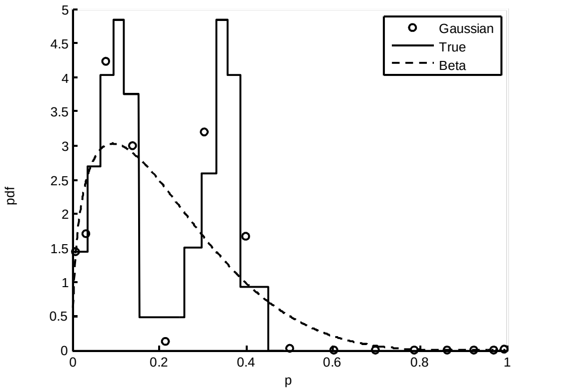
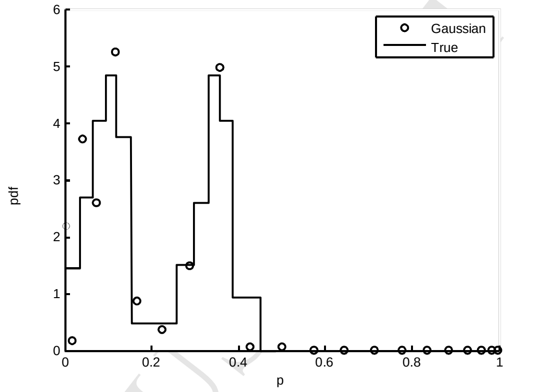


Figure 3b: Approximation of Piecewise Linear Distribution



Next, we evaluate the predictive performance of our GB/NBD model. We simulate one million customers' purchases within a time range $(0, t)$. We split customers' purchases (X) into two parts (X_1, X_2), where X_1 and X_2 denote the purchases within the time ranges $(0, t/2)$ and $(t/2, t)$, respectively. We use X_1 to estimate the *k+2* parameters for the GB/NBD model and then use the parameters estimated to predict the same customer's future purchase (X'_2) within $(t/2, t)$. The predicted results X'_2 are compared with the customers' real purchases in the time range of $(t/2, t)$.

We follow the same procedure and obtain the predictive results for the BB/NBD model. To compare the predictive performance of GB/NBD and BB/NBD, we conduct the simulation for three different selection rates' distributions, piecewise linear and truncated normal distributions with mean 0.3 and 0.5. The standard deviations for the truncated normal distributions are kept the same at 0.1. The pdf curves of these distributions are shown in Figures 2a to 3b. The resulting RMSE values are presented in Table 6, where the numbers in the brackets are the orders of the Gaussian quadrature used in the simulation. Under all the three selection rates' distributions, the RMSE value of

[2] The main idea of Gaussian quadrature is polynomial approximation. We need a higher order polynomial function to approximate a non-smooth curve.

GB/NBD is smaller than that of BB/NBD, which indicates a better prediction performance by GB/NBD. For piecewise linear case, the RMSE by GB/NBD of using a 21 order Gaussian quadrature (6.5640) is smaller than that of using a 15 order Gaussian quadrature (6.6030), and that is to be expected. This is because a higher order Gaussian quadrature leads to a better approximation, and consequently better prediction results.

Table 6 RMSE of the Prediction Performance

	Piecewise linear	**Normal (0.3, 0.1^2)**	**Normal (0.5, 0.1^2)**
BB/NBD	7.1567	8.7461	16.8484
GB/NBD	6.5640 (21) 6.6030 (15)	7.9641(9)	14.5333(9)

6. Conclusions and Discussions

In this paper, we propose a stochastic model to predict customers' future spending based on their historical purchase transactions, without knowing customers' demographic information. The heterogeneity of customers is modeled using an arbitrary generalized distribution. In doing so, we extend the classical BB/NBD model to a more realistic model, which is applicable in O2O market, where a distinct feature is that the service provided by businesses are diverse, and thus the customers are highly diversified geographically and demographically.

We acquire a dataset from Ctrip.com, one of the leading O2O providers, to test our mode along with the traditional model. Overall, the proposed approach has both better in-sample and out-of-sample predictions. We further conduct a simulation study to validate all the underlying steps.

The proposed GB/NBD model can be implemented in any situation where BB/NBD can be used. The model proposed is especially suitable in online O2O contexts, where customers are much more diverted and the transaction volumes are much higher. As our simulation results show, when customers' selection rates of the focal brand are distributed by very complicate distributions (such as piecewise linear), the traditional

BB/NBD fails in terms of both fitness and prediction accuracy, where the GB/NBD model performs well. Note that by increasing the order of the Gaussian quadrature, we always can get satisfactory results from the GB/NBD model.

It is worth mentioning that even though we introduce more parameters into the model than BB/NBD does, the optimization process for the simulation of a group of one million customers can still finish in minutes on a modern computer (Intel i7). The parameter estimation part is not computationally challenging for the GB/NBD model.

The proposed model is not without limitation. The current model doesn't include customer's demographic or preference information. As we know, each customer's demographic or preference information is very difficult/costly to get. But in some situations when we have these individual level information, it will be very interesting to study how to incorporate them into the model, and to improve the prediction at each customer's level. On the other hand, for a O2O business, to effectively manage and allocate capacity, the aggregated, subcategory/brand level prediction is more relevant than at individual level prediction. For that purpose, the proposed GB/NBD model suffices.

References

Abe, M. 2009. "Counting your customers" one by one: A hierarchical Bayes extension to the Pareto/NBD model. Marketing Science, **28**(3), 541-553.

Abramowitz, M., I. A. Stegun (eds). 1972. *Handbook of Mathematical Functions,* Dover Publications (New York).

Chatfield, C., G. J. Goodhardt. 1970. The Beta-Binomial model for customer purchasing behavior. *J. Roy. Statist. Soc. Ser. C.* **19**(3), 240-250.

Chen, X., Wang, X., X. Jiang, 2016. The impact of power structure on the retail service supply chain with an O2O mixed channel. *Journal of the Operational Research Society*, **67**(2), 294-301.

Chen, Y. C., Hsieh, H. C., & Lin, H. C. 2013. Improved precision recommendation scheme by BPNN algorithm in O2O commerce. In e-Business Engineering (ICEBE), 2013 IEEE 10th International Conference on (pp. 324-328). IEEE.

Choi, T. M., Hui, C. L., Liu, N., Ng, S. F., Y. Yu, 2014. Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, **59**, 84-92.

Dunn, R., S. Reader, N. Wrigley. 1983. An investigation of the assumptions of the NBD model as applied to purchasing at individual stores. *J. Roy. Statist. Soc. Ser. C.* **32**(3), 249-259.

Du, Y., Y. Tang, 2014. Study on the Development of O2O E-commerce Platform of China from the Perspective of Offline Service Quality. *International Journal of Business and Social Science*, **5**(4).

Fader, P. S., B. G. S. Hardie. 2000. A note on modeling underreported Poisson counts. *J. Appl. Statist.* **27**(8), 953-964.

Fader, P. S., B. G. S. Hardie. 2001. Forecasting repeat sales at CDNOW: A case study. Interfaces, **31**(3_supplement), S94-S107.

Fader, P. S., Hardie, B. G., K. L. Lee, 2005. "Counting your customers" the easy way: An alternative to the Pareto/NBD model. Marketing science, **24**(2), 275-284.

Fader, P. S., B. G. S. Hardie. 2005. The value of simple models in new product forecasting and customer-base analysis. Applied Stochastic models in business and industry, **21**(4-5), 461-474.

Fader, P. S., B. G. S. Hardie. 2010. Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Sci.* **29**(1), 85-93.

Fader, P. S., D. C. Schmittlein. 1993. Excess behavioral loyalty for high-share brands: Deviations from the Dirichletmodel for repeat purchasing. *J. Marketing Res.* **30**(4), 478-493.

Goldstein, D. G., G. Gigerenzer, 2009. Fast and frugal forecasting. International Journal of Forecasting, **25**(4), 760-772.

Jeuland, A. P., F. M. Bass, G. P. Wright. 1980. A multibrand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes. *Oper. Res.* **28**(2), 255-277.

Morrison, D. G., D. C. Schmittlein. 1981. Predicting future random events based on past performance. *Management Sci.* **27**(9), 1006-1023.

Morrison, D. G., D. C. Schmittlein. 1988. Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *J. Bus. Econ. Statist.* **6**(2), 145-159.

Press, W. H., S., Teukolsky, W., Vetterling, B., Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

Ren, S., Choi, T. M., N. Liu, 2015. Fashion sales forecasting with a panel data-based particle-filter model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* **45**(3), 411-421.

Robbins, H. 1977. Prediction and Estimation for the Compound Poisson Distribution, *Proc. National Acad. Sci.* USA, Vol. **74** (1977), pp. 2670-2671.

Schmittlein, D. C., A. C. Bemmaor, D. G. Morrison. 1985. Why does the NBD model work? Robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Sci.* **4**(3), 255-266.

Stoer, J., R. Bulirsch. 2002, Introduction to Numerical Analysis (3rd ed.), Springer, ISBN 978-0-387-95452-3

Tsai, T. M., Yang, P. C., W. N. Wang, 2013. Pilot study toward realizing social effect in O2O commerce services. In International Conference on Social Informatics (pp. 268-273). Springer International Publishing.

Winkelmann, R. 2008. *Econometric Analysis of Count Data*, 5th ed. Springer, Berlin.

Xiao, S., M. Dong, 2015. Hidden semi-Markov model-based reputation management system for online to offline (O2O) e-commerce markets. *Decision Support Systems*, **77**, 87-99.

Zheng, Z., P. Fader, B.Padmanabhan. 2012. From business intelligence to competitive intelligence: Inferring competitive measures using augmented site-centric data. *Inform. Systems Res*. **23**(3), 698-720.

Appendix

The marginal distributions:

- $P(N|X)$
  - $h(\lambda, p|X = x)$

$$= \frac{P(X = x|\lambda, p)f(\lambda)g(p)}{P(X = x)} = \frac{p^x e^{-\lambda(p+\alpha)}\lambda^{r+x-1}g(p)}{\Gamma(x + r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}$$

  - $P(N = n|X = x, \lambda, p) = \begin{cases} \frac{[\lambda(1-p)]^{n-x}e^{-\lambda(1-p)}}{(n-x)!} & n \geq x \\ 0 & otherwise \end{cases}$
  - $P(N = n|X = x)$

$$= \int\int P(N = n|X = x, \lambda, p)h(\lambda, p|X = x)d\lambda dp$$

$$= \int\int \frac{[\lambda(1-p)]^{n-x}e^{-\lambda(1-p)}}{(n-x)!}\frac{p^x e^{-\lambda(p+\alpha)}\lambda^{r+x-1}g(p)}{\Gamma(x + r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}d\lambda dp$$

$$= \frac{1}{(n-x)!\,\Gamma(x + r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}\int_p p^x(1 - p)^{n-x}g(p)\int_\lambda \lambda^{n+r-1}e^{-\lambda(1+\alpha)}d\lambda dp$$

$$= \frac{\Gamma(n + r)}{(1+\alpha)^{n+r}(n-x)!\,\Gamma(x + r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}\int_p p^x(1-p)^{n-x}g(p)\,dp$$

$$= \frac{\Gamma(n + r)\sum_{i=1}^{k}p_i^x(1-p_i)^{n-x}g(p_i)w(i)}{(1+\alpha)^{n+r}(n-x)!\,\Gamma(x + r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}$$

  - $E(N|X = x) = \sum_{n=x}^{\infty} nP(N = n|X = x)$

$$= \sum_{n=x}^{\infty}\frac{n\Gamma(n + r)\sum_{i=1}^{k}p_i^x(1-p_i)^{n-x}g(p_i)w(i)}{(1+\alpha)^{n+r}(n-x)!\,\Gamma(x + r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}$$

$$= \frac{1}{\sum_{i=1}^{k}\frac{p_i^x}{(\alpha + p_i)^{x+r}}g(p_i)w(i)}\sum_{i=1}^{k}p_i^x g(p_i)w(i)\sum_{n=x}^{\infty}\frac{n\Gamma(n + r)(1-p_i)^{n-x}}{\Gamma(x + r)(1+\alpha)^{n+r}(n-x)!}$$

$$= \frac{1}{\sum_{i=1}^{k} \frac{p_i^x}{(\alpha+p_i)^{x+r}} g(p_i)w(i)} \sum_{i=1}^{k} p_i^x g(p_i) w(i) \left(\frac{1}{\alpha+p_i}\right)^{r+x} \left(\frac{(r+x)(1-p_i)}{\alpha+p_i} + x\right)$$

$$= \frac{\sum_{i=1}^{k} \frac{x(1+\alpha)+r(1-p_i)}{\alpha+p_i} \left(\frac{p_i}{\alpha+p_i}\right)^x \left(\frac{1}{\alpha+p_i}\right)^r g(p_i)w(i)}{\sum_{i=1}^{k} \left(\frac{p_i}{\alpha+p_i}\right)^x \left(\frac{1}{\alpha+p_i}\right)^r g(p_i)w(i)}$$

- $\sum_{n=x}^{\infty} \frac{n\Gamma(n+r)(1-p_i)^{n-x}}{\Gamma(x+r)(1+\alpha)^{n+r}(n-x)!} =$
  $\sum_{n'=0}^{\infty} \frac{n'\Gamma(n'+x+r)(1-p_i)^{n'}}{\Gamma(x+r)(1+\alpha)^{n'+x+r}n'!} + x\sum_{n'=0}^{\infty} \frac{\Gamma(n'+x+r)(1-p_i)^{n'}}{\Gamma(x+r)(1+\alpha)^{n'+x+r}n'!}$

Let $x + r = r'$

$$= \sum_{n'=0}^{\infty} \frac{n'\Gamma(n'+r')(1-p_i)^{n'}}{\Gamma(r')(1+\alpha)^{n'+r'}n'!} + x\sum_{n'=0}^{\infty} \frac{\Gamma(n'+r')(1-p_i)^{n'}}{\Gamma(r')(1+\alpha)^{n'+r'}n'!}$$

$$= \sum_{n'=0}^{\infty} \frac{n'\Gamma(n'+r')(1-p_i)^{n'}}{\Gamma(r')(1+\alpha)^{n'+r'}n'!} + x\sum_{n'=0}^{\infty} \frac{\Gamma(n'+r')(1-p_i)^{n'}}{\Gamma(r')(1+\alpha)^{n'+r'}n'!}$$

$$= \left(\frac{1+\alpha'}{\alpha'(1+\alpha)}\right)^{r'} \left[\sum_{n'=0}^{\infty} \frac{n'\Gamma(n'+r')}{\Gamma(r')n'!} \left(\frac{\alpha'}{\alpha'+1}\right)^{r'} \left(\frac{1}{1+\frac{\alpha+p_i}{1-p_i}}\right)^{n'} + x\sum_{n'=0}^{\infty} \frac{\Gamma(n'+r')}{\Gamma(r')n'!} \left(\frac{\alpha'}{\alpha'+1}\right)^{r'} \left(\frac{1}{1+\frac{\alpha+p_i}{1-p_i}}\right)^{n'}\right]$$

$$= \left(\frac{1+\alpha'}{\alpha'(1+\alpha)}\right)^{r'} \left(\frac{r'}{\alpha'} + x\right) = \left(\frac{1+\frac{\alpha+p_i}{1-p_i}}{\frac{\alpha+p_i}{1-p_i}(1+\alpha)}\right)^{r'} \left(\frac{r'}{\frac{\alpha+p_i}{1-p_i}} + x\right)$$

$$= \left(\frac{1}{\alpha+p_i}\right)^{r'} \left(\frac{r'(1-p_i)}{\alpha+p_i} + x\right)$$

- $g(p|X)$

$$g(p|X=x)=\int_0^\infty h(\lambda,p|X=x)d\lambda=\int_0^\infty \frac{p^x e^{-\lambda(p+\alpha)}\lambda^{r+x-1}g(p)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}d\lambda$$
$$=\frac{p^x g(p)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}\int_0^\infty e^{-\lambda(p+\alpha)}\lambda^{r+x-1}d\lambda$$
$$=\frac{p^x g(p)}{(\mathrm{p}+\alpha)^{\mathrm{r+x}}\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}$$

$$E(p|X=x)=\int_0^1 pg(p|X=x)dp$$
$$=\int_0^1 \frac{p^{x+1}g(p)}{(\mathrm{p}+\alpha)^{\mathrm{r+x}}\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}dp$$
$$=\frac{\sum_{i=1}^{k}\frac{p_i^{x+1}}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}{\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}$$

- $f(\lambda|X)$

$$f(\lambda|X=x)=\int_0^1 h(\lambda,p|X=x)dp=\int_0^1 \frac{p^x e^{-\lambda(p+\alpha)}\lambda^{r+x-1}g(p)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}dp$$
$$=\frac{e^{-\lambda\alpha}\lambda^{r+x-1}}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}\int_0^1 e^{-\lambda p}p^x g(p)dp$$
$$=\frac{e^{-\lambda\alpha}\lambda^{r+x-1}\sum_{i=1}^{k}e^{-\lambda p_i}p_i^x g(p_i)w(i)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}$$

$$E(\lambda|X=x)=\int_0^\infty \lambda f(\lambda|X=x)d\lambda=\int_0^\infty \frac{\lambda e^{-\lambda\alpha}\lambda^{r+x-1}\sum_{i=1}^{k}e^{-\lambda p_i}p_i^x g(p_i)w(i)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}d\lambda$$
$$=\frac{\sum_{i=1}^{k}p_i^x g(p_i)w(i)\int_0^\infty e^{-\lambda(\alpha+p_i)}\lambda^{r+x}d\lambda}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}$$
$$=\frac{\Gamma(r+x+1)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r+1}}g(p_i)w(i)}{\Gamma(x+r)\sum_{i=1}^{k}\frac{p_i^x}{(\alpha+p_i)^{x+r}}g(p_i)w(i)}$$