Modelling Passive Forever Churn via Bayesian Survival Analysis

Gavin Steininger
Phoenix Labs[a)]

This paper presents an approach to modelling passive forever churn (i.e., the probability that a user never returns to a game that does not require them to cancel it). The approach is based on parametric mixture models (Weibull, Gamma, and Log-normal) for return times. The model and data are inverted using Bayesian methods (MCMC and DIC) to get parameter estimates, uncertainties, as well as determine the return time distribution for retained users. The inversion scheme is tested on three groups of simulated data sets and one observed data set. The simulated data are generated with each of the parametric models. Each data set is censored to six time horizons, creating 18 data sets. All data sets are inverted with all three parametric models and the DIC is used to select the return time distribution. For all data sets the true return time distribution (i.e., the one that is used to simulate the data) has the best DIC value; for 16 inversions the true return time distribution is found to be significantly better than the other options. For the observed data set inversion, the scheme is able to accurately estimate the % of users that did return (before the game transitioned into open beta) to given 14 days of observations.

PACS numbers:

I. Free to play games and Passive Churn

Free to play gaming is a large industry world wide. In free to play gaming it is common for games to be thought of as a service instead of a product. As with many services reducing churn is valuable. In free to play gaming defining churn can be difficult as there is is no definitive action taken by users to indicate they are churning, instead users simply stop returning to the game; i.e., it is never clear if a user has actually churned or simply reduced their login velocity. This passive churn (PC) is therefor more difficult to model than active churn (i.e., a user notifies a service of their cancellation).

A popular strategy to define churn is to give users of a product a specific amount of time to return and define all users that have not returned in that time as having churned.[1] While this approach has obvious value, it can be improved upon as it discards a lot of information. In particular, it does not consider the rate of user return or the returns after the time interval. In-addition, the fixed interval method requires that the full time interval must be observed before conclusions on churn can be made.

More recently[2,3] churn has been modelled by predicting the time between player actions using survival analysis. Note that the time until churn is not modelled instead the survival analysis is used to model the time between a users exiting a service as part of their normal flow and returning to it again (e.g., log out of a game and log in again). Thus this method is used to model user activity velocity. Intuitively as user activity velocity increases user churn decreases and visa versa.

Here a Bayesian survival analysis approach is proposed and tested on both simulated and observed data. The proposed approach consists of modelling the inter event time using parametric mixture models; parameters are estimated using Markov chain Monte Carlo[4–6] and parameterization selection is conducted using the deviance information criterion.[4,5]

TABLE I. Sample Survival Time Data (**d**)

Observation Number	Time (**t**)	Status (**e**)
1	1.7	1
2	21.2	0
...	...	...
n	3.2	1

II. Background theory

This section briefly reviews the relevant parts of survival analysis[7] and Bayesian inversion[4,5]. In particular the goal is to show how an observed set of player activity data can be used to generate a likelihood function, how that likelihood function can be augmented with prior information to create a posterior distribution, and finally how inference can be conducted using that posterior distribution.

Survival analysis is the study of time to a specified event.[7] Here the time to event is the time between logins (The time between last log out and the next login) for returning users sessions or the time between last log out and now (the time the data is harvested) for churned users sessions. The event time is denoted **t**. In addition to the event time, the event status (**e**) is also recorded. The status indicates wether a user session is the last observed (i.e., the user has not returned). The status is either a 1 or 0; i.e., $e_i = 1$ indicates that the ith observation has returned at the recored time and $e_i = 0$ indicates that the ith observation has not yet returned. Collectively **t** and **e** are the data (**d**). For this method **d** should be of the form of Tab. I; i.e., each observation (1 to n) should have a time and an event. Thus all data is utilized no mater how recently a player's most recent session ended. More recent unreturned log outs simply provide less information that older ones.

[a)] Electronic mail: gavin.amw.steininger@gmail.com

To define the probability for a set survival data it is helpful to consider returning and non returning observations separately. Staring with the observed return times (i.e., $\mathbf{d}$ where $\mathbf{e}$ is 1). Let $f(t)$ be the distribution of user return times; i.e., $f(t = t_\star)$ is the probability that a user session has exactly $t_\star$ time between log out and the start of the next session. For the unreturned the users sessions note that they are considered censored not churned; this is because all that is know about unreturned users are that they have not yet returned not that they will never return. The probability of observing a censored data point (i.e., one index $\mathbf{d}$ where $\mathbf{e}$ is 0) is $S(t = t_\star) = 1 - F(t = t_\star)$ where $S()$ is called the survival function and $F()$ is the cumulative density function (i.e., $F(t) = \int_{-\infty}^{t} da f[a]$). Thus the probability of witnessing a censored observation at time $t_\star$ is one minus the probability witnessing it return by $t_\star$ and is $S(t = t_\star) = 1 - F(t = t_\star)$. There is still one more important function to define before the probability of the data set can be expressed; that is the hazard function (h[t]). The hazard function is the instantaneous probability of an event given that it has not yet occurred;[7] i.e.,

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log[S(t)]. \tag{1}$$

Thus for survival analysis with only right censored data (as is the case here) the probability of a set of observations can be expressed as

$$P(\mathbf{d}) = \Pi_{i=1}^{n} S(t_i) h(t_i)^{e_i}. \tag{2}$$

If an observation is censored the probability of the observation is given by $S(t)$ and if it is not censored it is given by $f(t)$.

Survival data, specifically the empirical survival function, is commonly visualized using Kaplan-Meier plots.[7] Figure 1 shows a Kaplan-Meier of three simulated data sets (their details are given in Sec. V). On Kaplan-Meier plots the "x" axis is normally time, and the "y" axis is the proportion of observations that have not yet had an event. The steps indicate one or more events happening at a time, the crosses show the times that observations become censored.

In Bayesian inference, model parameters are considered random variables; let $\mathbf{m}$ represent an arbitrary model. Using Bayes' rule, posterior probability density (PPD) can be written[4–6]

$$P(\mathbf{m}|\mathbf{d}) = \frac{\pi(\mathbf{m})\mathcal{L}(\mathbf{m})}{\mathcal{Z}}\,, \tag{3}$$

where $\pi(\mathbf{m})$ is the prior distribution of $\mathbf{m}$, $\mathcal{Z}$ is the total Bayesian evidence of the ensemble of models (i.e., the probability of the data integrated and summed over all possible parameter values and parameterizations), and $\mathcal{L}(\mathbf{m})$ is the likelihood of the parameter vector (i.e., the probability of the data given the model ($\mathcal{L}(\mathbf{m}) = P[\mathbf{d}]$); note that the likelihood is considered a function of $\mathbf{m}$ as the observed data [$\mathbf{d}$] are considered known when the posterior distribution is evaluated).

The likelihood function contains all of the data information in an inverse (model estimation) problem, however that is not all the information about a problem. The prior distribution contains all the *a priori* information about the inverse problem. For example previous observations of a product (or similar products) may show that median return time is x days or it may be the case that return times in excess of 365 days are effectively equivalent to return times of 3,650 days and therefore there is no point in precise estimation of parameters above that value. In either case it helpful to incorporate this prior information into a problem. Thus the posterior probability distribution contains both data and prior information.

The posterior distributions used in this work are not analytically tractable; so they must be numerically approximated. The approximation is conducted using Markov chain Monte Carlo (MCMC).[4–6] The MCMC process is implemented with parallel tempering[8] (PT) and parameter rotation[9]. MCMC consists of taking a random walk through the parameter space. At each step of the random walk either the process will stand still (repeat the current location) or move to a new location. The record of locations (including repeats) is used as an approximation for the posterior distribution. The adoption of PT means that a population of Markov chains are sampled each at a different temperature (β). A tempered Markov chain has the standard likelihood raised to the power of $1/\beta$; i.e., the tempered likelihood is $\mathcal{L}^{1/\beta}(\mathbf{m})$. Higher temperature chains are more likely to accept lower likelihood models and lower temperature chains are less likely. Consequently the higher temperature chains sample the posterior more broadly while the lower temperature chains focus on higher probability spaces. The chains periodically interact (swap temperatures); this is done such that on average the lower temperature chains are assigned higher likelihood models. Only samples from the chain with a temperature of 1 are used to describe the posterior distribution.

In more detail let $\mathbf{m}$ be the current Markov chain state for a chain with sampling β temperature and $Q(\mathbf{m}'|\mathbf{m})$ be the proposal distribution by which a new state $\mathbf{m}'$ is generated. The proposed model represents a perturbation of the parameters of $\mathbf{m}$. The proposed state $\mathbf{m}'$ is accepted with probability

$$a = \min\left[1, \frac{\pi(\mathbf{m}')}{\pi(\mathbf{m})} \frac{\mathcal{L}^{1/\beta}(\mathbf{m}')}{\mathcal{L}^{1/\beta}(\mathbf{m})} \frac{Q(\mathbf{m}|\mathbf{m}')}{Q(\mathbf{m}'|\mathbf{m})}\right]. \tag{4}$$

The different chains of the population swap models periodically (here this event was triggered after the whole population of chains makes a MCMC step with a probability of 0.25). Pairs of chains are selected at random; the probability that two chains (i and j) swap temperature is

$$a = \min\left[1, \left\{\frac{\mathcal{L}(\mathbf{m}_i)}{\mathcal{L}(\mathbf{m}_j)}\right\}^{1/\beta_j - 1/\beta_i}\right]. \tag{5}$$

Only the steps from the chain with $\beta = 1$ are recorded.

Returning to the proposal distribution $Q(\mathbf{m}'|\mathbf{m})$, parameter rotation is incorporated into the sampling scheme so that Q does not in general have to perturb

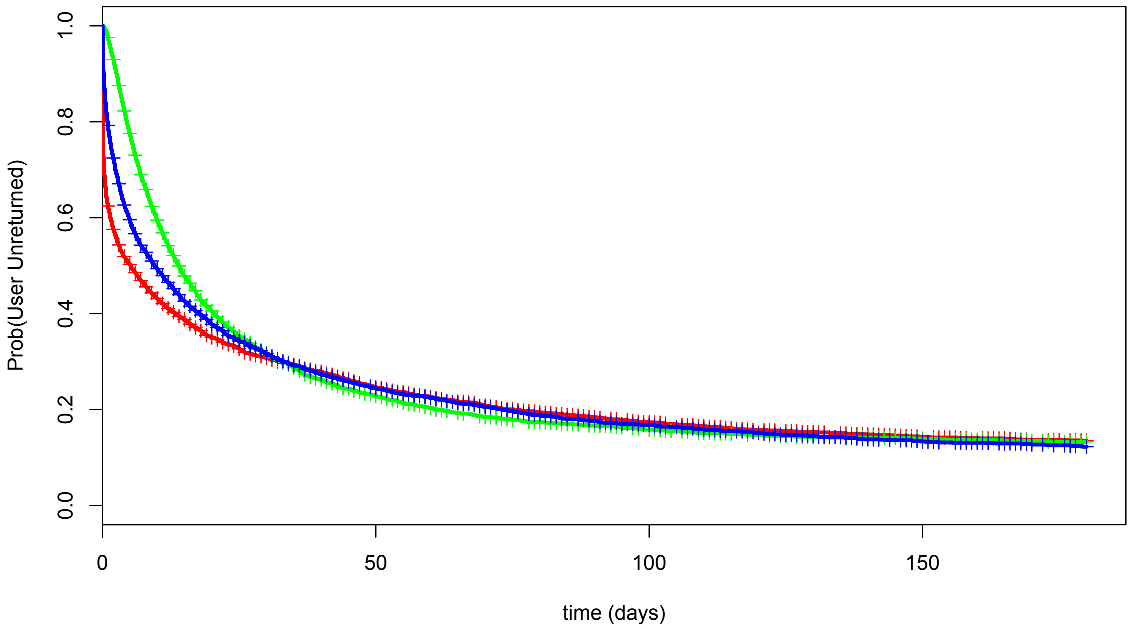

FIG. 1. Kaplan-Meier plot of the $\mathbf{d}_{\mathrm{W365}}$ (green line) and $\mathbf{d}_{\gamma 365}$ (red line), and $\mathbf{d}_{\text{Log-N365}}$ (blue line). The crosses indicate times of censor events.

only one parameter at a time but instead proposes models that are offsets along the principle components of the poster distribution. This is accomplished by sampling in the rotated space $\tilde{\mathbf{m}} = U^T \mathbf{m}$ and then rotating back to the standard parameter space using $\mathbf{m} = U\tilde{\mathbf{m}}$. Where the matrix U is is the column eigenvector matrix of the posterior covariance matrix ($C_{\mathbf{m}}$),

$$C_{\mathbf{m}} = UWU^T, \tag{6}$$

and $W = \mathrm{diag}[w_i]$ is the eigenvalue matrix with w_i representing the parameter variance projected along the eigenvector $\mathbf{u}_i$. $C_{\mathbf{m}}$ is approximated by the covariance posterior samples of the posterior distribution. Note that this violates the assumptions of MCMC; however in practice the estimated model covariance matrix will converge much faster than the sampling as a whole and thus there will be little impact on the posterior samples.

Feature selection (i.e., choice of parameterization) can be conducted on posterior samples using the deviance information criterion (DIC).[10] The DIC is used here to differentiate between models defined by choice of event time distribution. The DIC is a measure of model support based on a trade-off of data fit versus model complexity (parsimony) similar to other more common model-selection criteria such as the Bayesian information criterion (BIC).[4–6] However, the DIC is calculated from the posterior samples (rather than from a point-estimate of the best-fit model in BIC), and consequently has the advantages that it accounts for the prior distribution, parameter correlations, and the general non-Gaussianity of the posterior probability density.[10] The DIC trades off the data fit of a characteristic model against a complexity term and is defined as

$$\mathrm{DIC} = D(\hat{\mathbf{m}}) + 2P_D, \tag{7}$$

where D is the posterior deviance

$$D(\mathbf{m}) = -2\log[\mathcal{L}(\mathbf{m})] \tag{8}$$

and $\hat{\mathbf{m}}$ is a central or characteristic model (here the maximum *a posteriori* model is user). The term P_D in the DIC is the effective number of focused parameters (i.e., parameters that are not marginalized out of the posterior prior to DIC calculation). The $P_D = \overline{D(\mathbf{m})} - D(\hat{\mathbf{m}})$ where $\overline{D(\mathbf{m})}$ is the mean of the posterior deviance.

III. Estimating forever Churn

Forever Churn (FC) is the probability that a user never returns to a product. A common assumption in survival analysis is that ultimately there are no survivors; i.e., $\lim_{t\to\infty} S(t) = 0$. Thus to incorporate the concept that a user will never return the survival function must be modified such that $\lim_{t\to\infty} S(t) = \alpha$ the probability that a user is forever churned. This type of modification is called a parametric mixture model and is more commonly used in modelling cure rates.[7] Let $S(t)$ be a standard survival function then

$$S_{\mathrm{FC}}(t) = \alpha + (1-\alpha)S(t) \tag{9}$$

is the forever churn survival function. Substituting $S_{\mathrm{FC}}(t)$ into Eq. 1 the forever churn hazard function can be found to be

$$h_{\mathrm{FC}}(t) = \frac{f(t)}{S(t) + \frac{\alpha}{(1-\alpha)}}. \tag{10}$$

Because of the predictive nature of estimating forever churn, the event time distribution $f(t)$ must be specified parametrically. In this work three distributions (Weibull, Gamma, and Log-Normal)[7] are used. The probability distribution function for the for the Weibull, Gamma, and Log-Normals distributions are

$$f_{\mathrm{W}}(t) = \lambda\kappa[\lambda t]^{\kappa-1}\exp(-[\lambda t]^{\kappa}), \tag{11}$$

$$f_{\gamma}(t) = \frac{1}{\Gamma(\phi)\theta^{\phi}} t^{\phi-1}\exp(-\frac{t}{\theta}), \tag{12}$$

and

$$f_{\text{Log-N}}(t) = \frac{1}{\sqrt{2\pi}\sigma t}\exp\left(-\frac{1}{2}\frac{[\log(t)-\mu]^2}{\sigma^2}\right), \tag{13}$$

respectivly. Their survival functions are

$$S_{\mathrm{W}}(t) = \exp(-[\lambda t]^{\kappa}), \tag{14}$$

$$S_{\gamma}(t) = 1 - \frac{1}{\Gamma(\phi)}\gamma(\phi, t/\theta), \tag{15}$$

where $\gamma(y, x)$ is the lower incomplete gamma function[7], and

$$S_{\text{Log-N}}(t) = 1 - \frac{1}{2}\mathrm{erfc}\left(-\frac{[\log(t)-\mu]^2}{\sigma^2}\right). \tag{16}$$

Thus using Eqs. 9 and 10 the forever churn survival and hazard function for Weibull distributed return times become

$$S_{\mathrm{FCW}}(t) = \alpha + (1-\alpha)\exp(-[\lambda t]^{\kappa}) \tag{17}$$

and

$$h_{\mathrm{FCW}}(t) = \frac{\lambda\kappa[\lambda t]^{\kappa-1}\exp(-[\lambda t]^{\kappa})}{\exp(-[\lambda t]^{\kappa}) + \frac{\alpha}{(1-\alpha)}}, \tag{18}$$

respectively. $S_{\mathrm{FCW}}(t)$ and $h_{\mathrm{FCW}}(t)$ can be substituted into Eq. 2 to create forever churn likelihood for Weibull distributed events. The equivalent substitutions can be used for both the Gamma and Log-Normal distributions to create their likelihood functions. In all three cases the likelihood function have three parameters α and two distribution specific parameters.

IV. Parameterization and Priors

In order to facilitate the creation of parameter bounds and prior distributions it is helpful to re parameterize the likelihood parameters (e.g., α, λ, and κ) as functions of unknown model parameters $\mathbf{m}$; i.e., $\star = f_{\mathrm{Link}\star}[g_{\star}(\mathbf{m})]$ where $\star$ is represents an arbitrary parameter, $f_{\mathrm{Link}\star}$ is a link function, and $g_{\star}$ is function of model parameters and observation attributes. This representation of the unknown parameters can easily incorporate the attributes of the individual observations within the population; i.e., the survival and hazard functions of the ith observation ($S_i[t]$ and $h_i[t]$) do not necessarily equal the corresponding functions for the jth observation. This plurality is accomplished by including other attributes of the observations in the functions. Specifically the $g_{\star}$ can be considered a vector function with one entry for each observation (i.e., $g_{\star i}$) and can use the known attributes of each observation. Commonly[7]and here the $g_{\star}$ are taken as linear functions (There is evidence that in-general non linear functions are better,[1] but such functions tend to be more domain specific and describing them would be outside the scope of this work). For linear models the vector $g_{\star}(\mathbf{m}) = A\mathbf{m}_{\star}$ where A is the matrix of observation attributes having one row per observation and one column per attribute / feature (for a linear problem A would be the sensitivity matrix[4,5]) and the full parameter vector $\mathbf{m}$ is the concatenation of all parameter sub sets $\mathbf{m}_{\star}$. There is no requirement that all likelihood parameters vary between observations; in particular in the case that a proportional hazards assumption is reasonable for a Weibull model only α and λ need to vary between observations.[7] Even if multiple likelihood parameters are allowed to vary between observations it is not necessary that they all share the same features. Here for simplicity, A is assumed to be a column vector of ones. In addition, this parameterization naturally restricts the parameters to their meaningful ranges through the use of link the functions (e.g., $\alpha_i = \mathrm{Probit}(g_{\alpha i}[\mathbf{m}])$). The parameterized Weibull model is $\alpha = \mathrm{Probit}(m_1)$, $\alpha = \exp(-m_2)$, and $\kappa = m_3$. The parameterized Gamma model is $\alpha = \mathrm{Probit}(m_1)$, $\theta = \exp(m_2)$, and $\phi = m_3$. And the parameterized Log-Normal model is $\alpha = \mathrm{Probit}(m_1)$, $\mu = m_2$, and $\sigma^2 = m_3$

The priors for the model parameters must be selected with some care. It may be tempting to create uniform bounds or other semi-arbitrary convent functions around each parameter; however, this strategy will not work. Consider a Weibull model where $\lambda = 0$ and $\kappa = 1$ are known. Further suppose that all members of the population are homogeneous thus only one λ must be estimated. In this case $1/\lambda$ would then be the expected return time thus a uniform prior bound on λ would concentrate the expected time to less that 1 days! To express the prior information in prior distributions variable transforms[11] are used.

For all three return time distributions the forever churn probability α, a uniform prior is assumed (i.e., there is no knowledge about the probability of forever churn). Thus the prior on the α feature parameter (m_1) is

$$\pi_{\alpha}(m_1) = \mathrm{Norm}(0, 1), \tag{19}$$

where Norm represent the Gaussian distribution. For the return time distribution specific parameters a similar process can be applied. Starting with the Weibull model, *a priori* it is unlikely that strong information exists. To reflect this lack of knowledge let the median return time $(\log[2])^{1/m_3}\exp(m_2)$ be uniformly distributed between 0

TABLE II. True values of parameters for simulated data (**d**)

Simulation Description	data set	Parameters
Weibull	$\mathbf{d}_{\mathrm{W}}$	$\kappa = 0.5$
		$\lambda = 1/14$ days
		$\alpha = 10\%$
Gamma	$\mathbf{d}_{\gamma}$	$\phi = 0.2$
		$\theta = 140$ days
		$\alpha = 10\%$
Log-Normal	$\mathbf{d}_{\mathrm{Log\text{-}N}}$	$\sigma^2 = 1.791759 \log(\mathrm{days})^2$
		$\mu = 1.540446 \log(\mathrm{days})$
		$\alpha = 10\%$

and 180 days. Also to reflect that, in general, the risk of return decreases over time let κ be uniformly distributed between 0 and 1.Thus using a bi-variate variable transform the prior distribution of m_2 given m_3

$$\pi_\lambda(m_2|m_3) = \frac{(\log[2])^{1/m_3}}{180}\exp(-m_2), \tag{20}$$

where

$$\begin{array}{ll} m_2 \geq -\log(180) & \text{or} \\ m_3 \leq \log\left[\log(2)\right] / \left[m_2 + \log(180)\right] & \end{array} \tag{21}$$

and zero otherwise. Also $\pi_\kappa(m_3) = \mathrm{unif}(0,1)$.

The prior distribution for the log-Normal model is also found by assuming that the median return time is uniformly uncertain between 0 and 180 days. In addition, because the median of a log-normal distribution is $\exp(\mu)$ and free of σ^2 it is necessary to assume that that $\sigma^2 \sim$ exponentially. Thus,

$$\pi_\lambda(m_2, m_3) = \frac{\exp(m_2 - m_3)}{180}, \tag{22}$$

where $m_2 \leq \log(180)$.

The prior distribution Gamma model is derived similarly. The main difference is that since there is no simple analytical function for median return time the expected return time ($\theta\phi$) is used instead. While there can be a large difference between the mean and median of skewed distribution because the use here is only to express ignorance of the model parameters the exchange of the two central statistics is not important. Thus $\theta\phi$ is assumed to be uniformly distributed between 0 and 180 days and $\phi \sim \mathrm{unif}(0,1)$ again to reflect that risk is expected to decrease. Thus, m_2 given m_3

$$\pi_\lambda(m_2|m_3) = \exp(m_2)/180, \tag{23}$$

where $m_3 \leq 180\exp(-m_2)$ and zero otherwise.

V. Simulation Studies

The inversion scheme is tested on three simulated groups of data sets; these groups are each generated with one of the parametric models (Weibull, Gamma, and Log-Normal). The true parameter values are chosen such that the expected return times are 28 days and the standard deviation of the return times are $\sqrt{3,920}$ days. The true values of the parameters for simulated data sets are given in Tab. II. For each group of data sets the same 3,650 simulated uncensored observations are censored to different maximum time horizons to create six data sets. Thus in total there are 18 data sets (six per group). The time horizons for the data sets in each group are 7, 14, 28, 90, 180, and 365 days. In order to mimic observed game data each data set is divided into 365, 10 member parts, indexed 1-365. Observations in ith part are censored if they are larger than then minimum of i (the part index) and the data set time horizon. For example in a data set with time horizon 90 observations in the 1st part are censored if they are greater than 1. Observations in the second part are censored if they are greater than 2, and so on until the 91st part. Observations in parts 91 to 365 are censored if they are greater than 90 (the data set horizon time). In addition, to the standard censoring in order to simulate the forever churn percentage each observation has an independent 10% (α for each of the three groups) chance of being censored minimum of i the part index and the data set time horizon.

The Kaplan-Meier plots[7] of the 3 simulated 365 day data sets sets are shown in Fig. 1. Data sets $\mathbf{d}_{\mathrm{W}}$ is displayed as green, data set $\mathbf{d}_{\gamma}$ is red, and $\mathbf{d}_{\mathrm{Log\text{-}N}}$ is blue. The crosses indicate the times that observations are censored. The regular cadence of the censoring events is a result of the partitioning used to create the data.

All 18 data sets are inverted with all three paramedic models. Thus a total of of 54 inversion are conducted using the MCMC scheme described in Sec. II. The DIC is used to evaluate which parameterization has the most support from the data. Table III gives the DIC values for the Weibull, Log-Normal, and Gamma simulated data sets. In all cases the true parametrization is preferred by the DIC; i.e., the true parametrization has the lowest DIC. Significance of differences in the DIC values are assessed using standard Bayes factor tables;[12] i.e., ΔDIC < 2 is not significant and ΔDIC > 10 is very strong support. For the Weibull simulated data sets the Weibull parameterization is not found to be significantly better than the Gamma parameterization in the 7 and 14 day data sets; it is found to be significantly better in the 28, 90, 180, and 364 day data sets. In all other cases the true model is found to have significantly more support from the data than the alternative models.

The inversions of the Gamma data have noticeably lower DIC values for all parameterizations. This is surprising as *a prior* it is expected that the best DIC values would be found for the cases when inverted and true models match. The apparent reason for the low DIC (high likelihood) inversions seems to be the large proportion of quick ($t < 1$) events in the Gamma data set.

Figure 2 shows the marginal posterior density of the α parameters of the simulated inversions for the winning (lowest DIC) parameterizations. Each row shows the result form the inversion of one of the data sets (Weibull, Gamma, and Log-Normal) and each column shows results from a different time horizon. As the true model had lowest DIC in all cases the densities are shown form

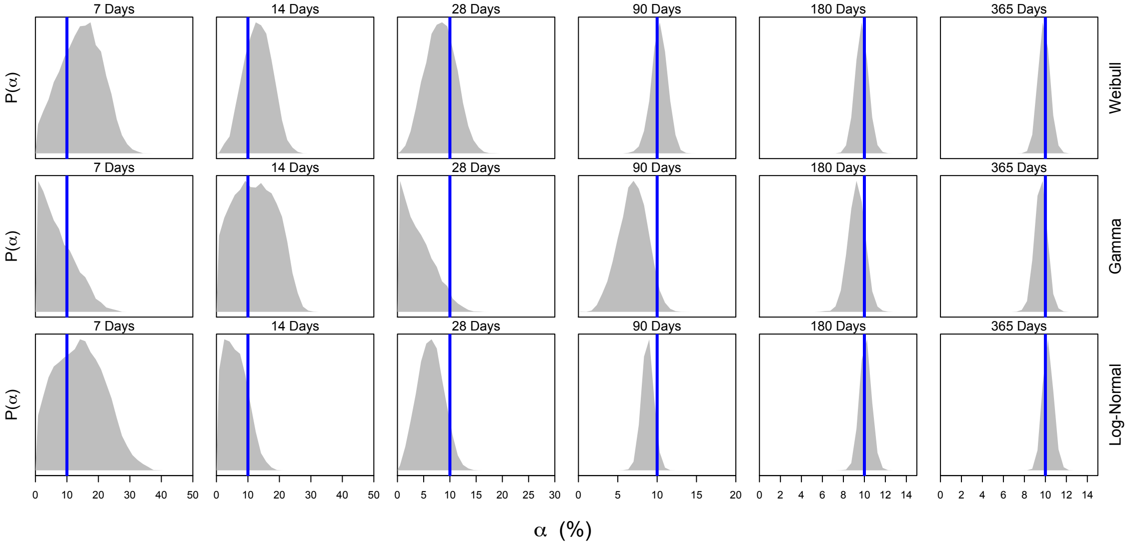

FIG. 2. Marginal posterior densities of the forever churn percentage α for the simulated inversions. Each row shows the result form the inversion of one of the data sets (row 1: Weibull, row 2: Gamma, row 3: Log-Normal). Each column shows results from a different time horizon. The vertical blue line indicates the true value. Note the x axis bounds differ between column to improve resolution.

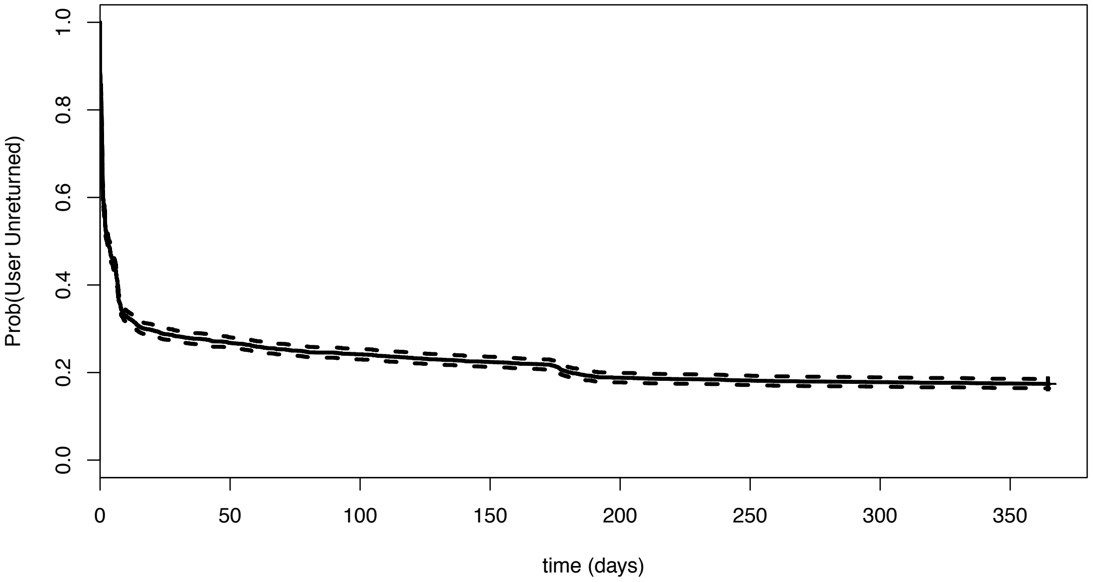

FIG. 3. Kaplan-Meier plot of the $\mathbf{d}_{\text{Daunt365}}$. The crosses indicate times of censor events and the dashed lines are indicate a 95% confidence bound.

the true parameterization. The vertical line indicates the true α value ($\alpha = 10\%$ for all simulated data sets). In all cases the true value is near the central mass of the posterior densities. The gamma inverted data is in general the least accurate and the Weibull is in general the most. The inversions with a 28 day time horizon are able to limit the poster range of FC to less than 15%. The inversions of data sets with time horizons of 90 days or more have well resolved poster densities centred near or on the true value.

Table IV gives the 2.5 and 97.5 posterior percentiles of α parameters of the winning (lowest DIC inversions) for each data set. In all cases the true value is within the bounds. Also the bounds generally shrink (95 % bound - 0.25% bound decreases) as the time horizons increase. However in several cases the (e.g., $\mathbf{d}_{\text{W14}}$ and $\mathbf{d}_{\text{W28}}$) the

TABLE III. The DIC values for the inversion of the simulated data sets. The columns indicate the assumed parameterization Weibull, Log-Normal, and Gamma. The rows indicate the inverted data set. The best (lowest) value for each data set is highlighted in bold. Statistically significantly best (ΔDIC > 10) values are underlined.

Data Set	Weibull	Log-Normal	Gamma
$\mathbf{d}_{\text{W7}}$	**9,696.122**	9,766.667	9,699.647
$\mathbf{d}_{\text{W14}}$	**13,134.255**	13,235.922	13,139.796
$\mathbf{d}_{\text{W28}}$	**17,039.807**	17,187.939	17,055.519
$\mathbf{d}_{\text{W90}}$	**22,037.123**	22,232.929	22,059.538
$\mathbf{d}_{\text{W180}}$	**23,584.058**	23,770.462	23,637.616
$\mathbf{d}_{\text{W365}}$	**23,922.285**	24,101.153	24,008.594
$\mathbf{d}_{\gamma 7}$	4,441.965	4,752.079	**4,416.549**
$\mathbf{d}_{\gamma 14}$	6,816.207	7,229.900	**6,775.124**
$\mathbf{d}_{\gamma 28}$	9,173.707	9,654.742	**9,137.870**
$\mathbf{d}_{\gamma 90}$	14,102.022	14,834.667	**14,026.442**
$\mathbf{d}_{\gamma 180}$	15,788.325	16,643.616	**15,706.391**
$\mathbf{d}_{\gamma 365}$	16,418.484	17,153.233	**16,339.646**
$\mathbf{d}_{\text{Log-N7}}$	9,050.803	**8,954.368**	8,964.477
$\mathbf{d}_{\text{Log-N14}}$	14,263.582	**14,169.804**	14,195.724
$\mathbf{d}_{\text{Log-N28}}$	19,718.789	**19,627.713**	19,674.905
$\mathbf{d}_{\text{Log-N90}}$	25,522.984	**25,371.625**	25,533.565
$\mathbf{d}_{\text{Log-N180}}$	26,593.123	**26,379.627**	26,639.120
$\mathbf{d}_{\text{Log-N365}}$	27,030.317	**26,741.268**	27,114.766

TABLE IV. The 2.5 and 97.5 posterior percentiles of α parameters of the simulated inversions for the true parameterizations.

Horizon	$\mathbf{d}_{\text{W}\star}$		$\mathbf{d}_{\text{Log-N}\star}$		$\mathbf{d}_{\gamma\star}$	
(days)	2.5 (%)	97.5 (%)	2.5 (%)	97.5 (%)	2.5 (%)	97.5 (%)
7	1.92	26.57	1.36	29.51	0.28	19.62
14	4.75	21.46	0.45	14.06	0.89	24.23
28	2.80	13.78	1.78	11.07	0.14	10.77
90	8.15	12.26	7.33	10.27	3.24	10.30
180	8.54	10.99	9.04	11.20	7.71	10.77
365	8.79	10.98	9.30	11.26	8.43	10.73

poster bounds of shorter time horizons do not span the bounds of longer time horizons. This is unexpected as it means that the local information at the end of the survival curve effects its forecast.

VI. Dauntless Data Inversion

TABLE V. The DIC values for the inversion of the observed Dauntless data sets. The columns indicate the assumed parameterization Weibull, Log-Normal, and Gamma. The rows indicate the inverted data set. The best (lowest) value for each data set is highlighted in bold. Statistically significantly best (ΔDIC > 10) values are underlined.

Data Set	Weibull	Log-Normal	Gamma
$\mathbf{d}_{\text{Daunt7}}$	6,707.824	7,036.155	**6,611.831**
$\mathbf{d}_{\text{Daunt14}}$	8,019.588	8,285.977	**7,835.913**
$\mathbf{d}_{\text{Daunt28}}$	9,409.964	9,708.785	**9,239.990**
$\mathbf{d}_{\text{Daunt90}}$	**11,315.602**	11,632.988	11,438.400
$\mathbf{d}_{\text{Daunt180}}$	**14,332.120**	14,508.708	14,736.985
$\mathbf{d}_{\text{Daunt365}}$	**15,090.318**	15,245.551	15,683.405

The inversion scheme is used on a group of observed data sets from the free to play PC game Dauntless. The data sets have 3,650 observations and similar to the simulated data sets they differ in that they are truncated by time horizon. For all observations the users last logout was recored on 2017-12-10; the observations are a random sample of users from that day. The data sets are denoted $\mathbf{d}_{\text{Daunt}\star}$ where $\star$ is particular time horizon (i.e., 7,14,28, 90, 180, and 365); e.g, $\mathbf{d}_{\text{Daunt7}}$ has observations from 2017-12-10 to 2018-12-17 and $\mathbf{d}_{\text{Daunt365}}$ has observations from 2017-12-10 to 2018-12-10. The Kaplan-Meier cure for the $\mathbf{d}_{\text{Daunt365}}$ data set is shown in Fig. 3. Two important factors of the observed data sets are the oscillation in hazard over the first few days and the *step* around 180 days. The oscillation is caused by time of day; users tend to have sessions (if they return at all) at the same time each day. Thus the low hazard times are 12 hours offset from a users typical login time. The *step* is linked to the product changing from closed to open beta at that time (2018-05-24) and large number of apparently users returning. None of the parametric models used in the inversion can mimic either of features; they represent a systematic error.

As with the simulated inversions all three parameterizations are inverted for all data sets. The parameterization with the lowest DIC is preferred; the DIC values for the inversion are listed in Tab. V. In all cases the best parameterization is found to be significantly better than the other two for each data set. Unlike the simulated inversions the preferred parameterization changes with the time horizon; for the 7,14, and 28 day data sets the Gamma model is preferred and for the 90, 180, and 365 day data sets the Weibull parameterization is best. Thus it is likely that the true return-time distribution for the Dauntless data sets is not one of the assumed distributions. In particular, for the 180 and 365 day inversions the game open beta *step* seems to be effecting the estimates.

Figure 4 shows the posterior marginal distributions of α for the Dauntless data inversions. The marginal posterior distributions are centred $\sim$21% for the 7, 14,28, and 90 day horizon data sets for the inversions; $\sim$21% is approximately the percentage of users that had returned before the start of open beta. The 180, and 365 day inversions are centred near $\sim$ 15% and $\sim$ 18%, respectively. The posterior uncertainty of α for the 14 and 28 day inversions is quite small; $\sim$3.16% and $\sim$2.89%, respectively. Thus for the Dauntless data 14 days is sufficient to have a well resolved estimate of the passive forever churn. The anomalous α estimate for the 180 day inversion seems to be a result of the sharp increase in the hazard function near that time (the start of the open beta).

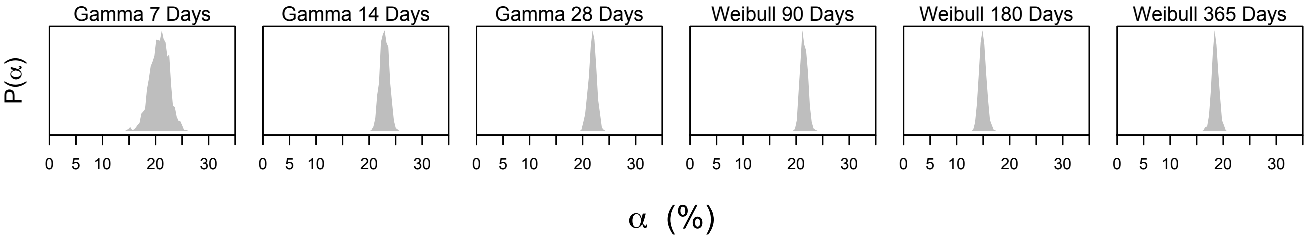

FIG. 4. Marginal posterior densities of the forever churn percentage α for the observed Dauntless data inversions. Each column shows results from a different time horizon.

VII. Conclusion

An approach for estimating passive forever churn in free to play games is proposed and tested on both simulated and observed data. The approach consists of using Bayesian inverse techniques (Markov chain Monte Carlo and deviance information criterion) to estimate the parameters of three parametric mixture models as well as choose which of the models is most likely to represent the data.

For all simulated data sets the inversion scheme is able to correctly choose the true parameterization (the DIC of the true parameterization is the lowest); in 16 of the 18 inversions the true parameterization was found to have significantly more support from the data that the other two parameterizations (only the Weibull parameterization for data sets with 7 or 14 day horizons were not found to be significantly better). The posterior marginal distributions of the passive forever churn percentages (α) are well resolved for the Weibull and Log-Normal simulated data inversions with 28 days (the expected return time) of uncensored event times. The Gamma simulated data inversion only becomes well resolved with 90 days ($\sim 3\times$ the expected return time) of uncensored event times. The posterior marginal densities of α not only shrink with the introduction of data but also shift central mass; thus, local trends in the survival curve can effect the forecasting of it and consequently the passive forever churn estimates. This vulnerability is the largest weakness of this methodology, but its importance should not be overstated as for most of the simulated data inversions the phenomena was not present. In addition, the true value of α is within the 95% central credibility interval of the posterior marginal distributions.

For the observed Dauntless data inversion the approach results in early data (7-28 days horizons) being classified as Gamma distributed and later data (90-365 days horizons) being classified as Weibull distributed. The marginal posterior distribution of α for the 14 days inversion is not meaningfully distinct from the 90 day inversion; in both cases α is $\sim 21\%$. This value is approximately the percent of users that had not retuned by the start of open beta.

In most practical applications it will be necessary to utilize the full flexibility of $g_{\star i}(\mathbf{m})$ from Sec. IV to allow the attributes of each observation to be incorporated into the likelihood function. Here that additional complexity was mostly excluded for the sake of clarity in introducing parametric mixture models, not because excluding observation attributes is somehow better.

Finally it should be noted that this approach can also be used to model other important events, e.g., first purchase. Thus it is possible to get estimates of the percentage of users that will never spend in a free to play game given a finite amount of time.

[1] M. Hassouna, A. Tarhin, T. Elyas, and M. S. AbouTrab, “Customer churn in mobile markets: A comparison of techniques”, International Business Research **8**, 224–237 (2015).

[2] A. G. Africa Perianez, Alain Saas and C. Magne, “Churn prediction in mobile social games: Towards a complete assessment using survival ensembles”, IEEE Int. Conf. Data Sci. Adv. Analytics 556–576 (2016).

[3] G. Steininger, “Bayesian business intelligence; how to define and model churn”, (2015), URL `https://bithebayesianway.wordpress.com/2015/08/03/how-to-define-and-model-churn/`.

[4] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 1–653 (Chapman, New York) (2004).

[5] D. G. T. Denison, C. C. Holmes, B. K. Mallick, and A. F. M. Smith, *Bayesian Methods for Nonlinear Classification and Regression*, 1–277 (Wiley, New York) (2002).

[6] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds., *Markov Chain Monte Carlo In Practice*, 1–486, Interdisciplinary Statistics (Chapman & Hall/CRC, Florida) (1996).

[7] D. Collett, *Modelling Survival Data in Medical Research*, 3rd edition (CRC press, 6000 broken sound parkway NW suite 300) (2015).

[8] D. J. Earl and M. W. Deem, “Parallel tempering: Theory, applications, and new perspectives”, Phys. Chem. Chem. Phys. **7**, 3910–3916 (2005).

[9] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*, 395–402, International Geophysics (Elsevier, Amsterdam) (2005).

[10] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, “Bayesian measures of model complexity and fit”, J. R. Statist. Soc. **64**, 583–639 (2002).

[11] G. Casella and R. Berger, *Statistical Inferance*, 1–84 (Thomson Learning Group, Oasific Grove) (2002).

[12] R. E. Kass and A. E. Raftery, “Bayes factors”, J. Am. Stat. Assoc. **90**, 773–795 (1995).