Prediction using Prospect Theory

David Walsh

December 16, 2013

Abstract

In this report, we consider prediction of an agent's preferences over risky monetary outcomes using Prospect Theory. We suppose that for a given agent we have data on previous prospects that the agent has accepted or declined. Based on this information, we would like to predict whether the agent will accept some new prospect X^*. This amounts to learning a value function v and probability weighting functions w^+, w^- for the agent in question, and using these functions to establish whether the prospect will look attractive to them.

In general, we do not expect to have sufficient data on a single agent to learn (v, w^+, w^-) from scratch. Instead, we assume that we have a population of agents with observed decision histories, on which the algorithm may first be trained. Each training agent has different values for (v, w^+, w^-). We assume an arbitrary parameterisation of these functions, controlled by some θ. Then, following the approach taken by Chajewska and Koller (2000) for prediction using Expected Utility Theory, we model θ as having some population-wide distribution P. Using P as a prior for the parameter value specific to the agent on whom we wish to make prediction, we may define a probability p^* that the agent is someone who would accept X^*.

We present an algorithm that uses a fully Bayesian framework to learn a posterior distribution for P, and hence a posterior distribution for p^*. This could be highly valuable in any competitive context where we expect an opponent to follow Prospect Theory, as it tells us which offers X^* they are likely to accept.

This approach differs from how PT is typically implemented in the literature. Most studies have been primarily descriptive and have sought to explain specific phenomena. Usually, heterogeneity in the functions (v, w^+, w^-) features as random effects terms and the model is fit using Maximum Likelihood.

Our algorithm is then made more robust by allowing a small proportion of agents to deviate from PT. That ensures that the posterior for P cannot be biased by a small number of training agents whose behaviour does not correlate with the predictions of PT. Further, if the agent on whom

we make prediction is not well described by PT, there is a risk that we conclude either $p^* = 0$ or 1 with great confidence. This refinement ensures a more conservative posterior distribution for p^* that avoids this.

1 Introduction

Prospect Theory [Tversky, Kahneman 1979, 1992] is a descriptive model for decision under risk that is commonly used across numerous applications. The key tenet of PT is that decisions are reference-specific, so that agents focus on immediate gains and losses. It also allows non-linear probability weighting, so that agents may be overly sensitive to tail events.

For a detailed exposition of the contexts to which PT has been successfully applied, see Barberis (2012). In general, PT has been applied most extensively in finance and insurance. For instance, many recent papers have used the over-weighting of tail probabilities associated with PT to explain why stocks with positively skewed returns (such those offered in an IPO) appear empirically to be overpriced [Barberis and Huang (2008), Boyer et al. (2010), Bali et al. (2011)]. Sydnor (2010) uses probability weighting to explain the purchase of insurance policies that charge a disaproportionately high premium to secure against very improbable disasters.

Other applications of PT include labor supply [Camerer et al. (1997), Koszegi, Rabin (2006)], gambling [Snowberg, Wolfers, 2010, Barberis, 2012] and retail [Heidhues, Koszegi (2012)].

In general, this algorithm is likely to be most useful in contexts where we are marketing prospects X^* to an individual customer and would like to know what offers they might accept. The package we are selling could be e.g. an insurance policy or a betting offer.

2 Model

A prospect is defined to be finite distribution over real-valued monetary outcomes. In particular, the prospect $(p_1 : x_1, ..., p_t : x_t)$ represents an opportunity to receive x_i with probability p_i.

We adopt specifically the Cumulative Prospect Theory model [Tverky, Kahneman 1992], under which an agent will accept a prosepct with $x_1 < ... < x_s \leq 0 < x_{s+1} < ... < x_t$ if

$$\sum_{r=1}^{s} \pi_i^- v(x_r) + \sum_{s+1}^{t} \pi_i^+ v(x_r) \geq 0 \tag{1}$$

$v : \mathtt{R} \to \mathtt{R}$ is a value function over outcomes specific to the agent, akin to utility in Expected Utility Theory. It is required to be continous and strictly increasing, with $v(0) = 0$. Generally it is assumed to be convex over losses and concave over gains, reflecting empirical loss aversion.

$\pi_i^{\pm}$ are decision weights, given by

$$\begin{aligned}
\pi_1^- &= w^-(p_1),\ \pi_t^+ = w^+(p_t) \\
\pi_r^- &= w^-(p_1 + ... + p_r) - w^-(p_1 + ... + p_{r-1}) && 2 \leq r \leq s \\
\pi_r^- &= w^+(p_r + ... + p_t) - w^+(p_{r+1} + ... + p_t) && s+1 \leq r \leq t-1
\end{aligned}$$

where $w^{\pm} : [0,1] \to [0,1]$ are probability weighting functions. These functions are required to be continuous and strictly increasing, with $w^{\pm}(0) = 0, w^{\pm}(1) = 1$. Often they are assumed to have an inverse S-shape: the larger curvature at the end points corresponds to the notion that people are more sensitive to differences in probabilities for events that have probabilities close to 0 or 1.

We suppose that we have training data on agents $i = 1, ..., n$, consisting of prospects $X^i = (X_1^i, ..., X_{m_i}^i)$ offered and binary labels $Z^i = (Z_1^i, ..., Z_{m_i}^i)$ indicating whether the prospect was accepted. Then, given a new agent for whom we have data $X' = (X_1', ..., X_{m'}'), Z' = (Z_1', ..., Z_{m'}')$ and a new prospect X^* that is offered to them, we seek to predict the agent's decision Z^*.

3 Parametric Specifications

We now briefly consider some parametric specifications for the functions (v, w^-, w^+). For a full discussion of the empirical evidence for alternative parameterisations, see Booij et al. (2009).

For the value function v, it is very common to use the power function

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\beta} & x \leq 0 \end{cases} \tag{2}$$

This corresponds to a CRRA utility function in Expected Utility Theory, and it was originally suggested by Tversky and Kahnemann (1992). It has strong empirical support [Wakker, 2008]. Alternative parameterisations include the exponential function, relating to CARA in EUT, and an amalgamated expo-power function (see Abdellaoui et al., 2007 for their properties).

There is greater variety in how people choose to model the weighting functions $w^{\pm}$. In general, it is considered desirable to be able to control the curvature and elevation of these functions separately. The curvature reflects an agent's ability to discern between mid-range probabilities, whereas the elevation reflects the agent's overall optimism in the case of w^+ or pessimism in the case of w^-.

Two possible parameterisations, introduced respectively by Goldstein and Einhorn (1987) and Prelec (1998), are given by

$$w(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma} \tag{3}$$

$$w(p) = exp\,\{-\delta(-log\,p)^\gamma\} \tag{4}$$

Here γ controls the curvature and δ controls the elevation.

The algorithm derived in this project works for any choice of parameterisation. In what follows, we suppose simply that (v, w^-, w^+) is controlled by some vector $\theta \in \Theta \subset \mathbb{R}^p$.

4 Training

Recall that an agent will accept a prospect X if (1) is satisfied for their particular (v, w^-, w^+). Viewing the expression in (1) as a function of θ, it follows that they will accept the prospect if and only if $\theta \in R_X$, where the set $R_X \subset \Theta$ is defined by some inequality $f(\theta) \geq 0$. For prospects $X = (X_1, ..., X_m)$ and responses $Z = (Z_1, ..., Z_m)$, set

$$\tilde{R}_{X_j,Z_j} = \begin{cases} R_{X_j} & Z_j = 1 \\ R^c_{X_j} & Z_j = 0 \end{cases} \qquad\qquad 1 \leq j \leq m$$

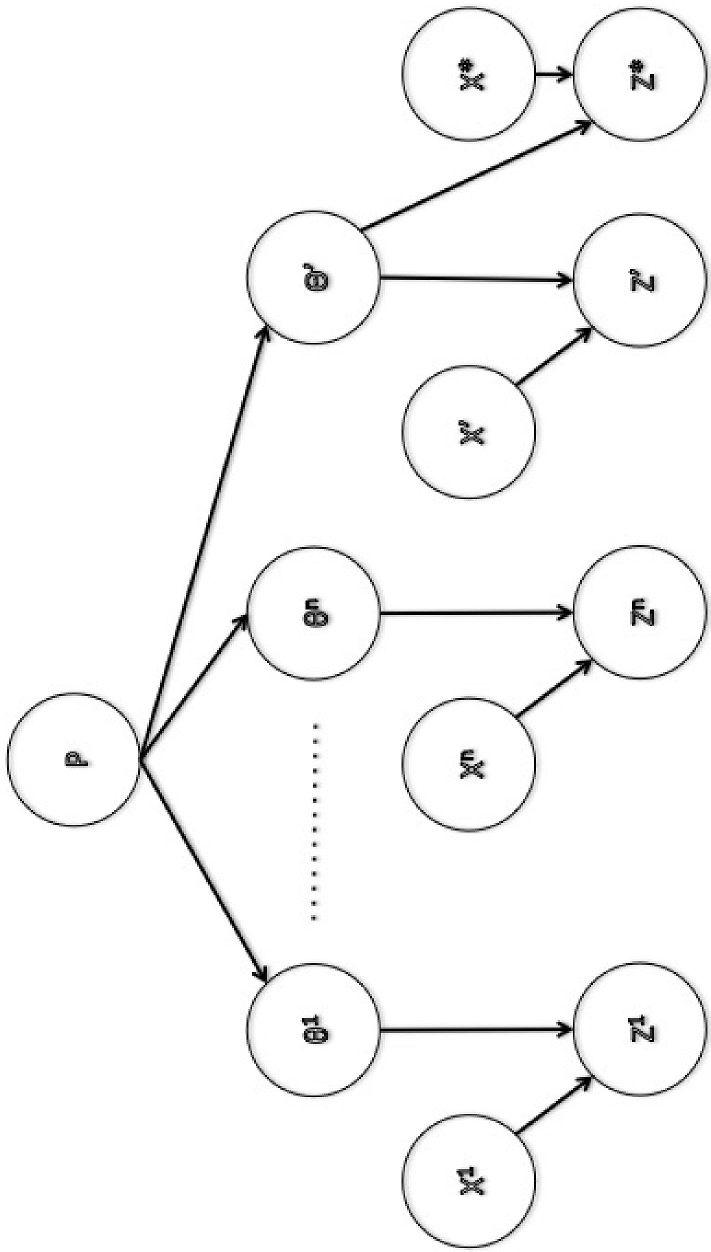
$$R_{X,Z} = \bigcap_{j=1}^{m} \tilde{R}_{X_j,Z_j}$$

The agent will give the exact sequence of responses Z to prospects X if and only if $\theta \in R_{X,Z}$. In other words, given data (X, Z), the likelihood of θ is

$$L(\theta) = 1(\theta \in R_{X,Z}) \tag{5}$$

Often, for a particular value of θ, we will need to check whether $\theta \in R_{X,Z}$ for some given prospects X and responses Z. It is worth noting that $R_{X,Z}$ is defined by m simultaneous inequalities, so that is reasonably straight-forward.

Now the focus of training the algorithm is to learn the population heterogeneity in the functions (v, w^-, w^+). For this parameterisation, that amounts to learning the population-wide distribution P of θ. We adopt a fully Bayesian framework to achieve this. As a prior for P, we use a Dirichlet Process with a suitably chosen measure α over Θ. The parameter values $\theta^1, ..., \theta^n$ for the agents in the training set then constitute an *iid* sample from P. From the decision histories of these agents, we may make inference on $\theta^1, ..., \theta^n$ using (5), and hence on P.



The posterior distribution of P, given the training agents' data $D = (X^1, Z^1, ..., X^n, Z^n)$ is derived by Gibbs Sampling.

For any $A \subset \Theta$,

$$p(\theta^i \in A | X^i, Z^i) = p(\theta^i \in A | \theta^i \in R_{X^i, Z^i}) \tag{6}$$

Hence the posterior for $\theta^i \mid P$ is given by

$$\therefore \theta^i \mid P, D \sim P|_{R_{X^i, Z^i}} \tag{7}$$

where $P|_A$ denotes the distribution of P conditioned to the set A. We can sample from this distribution by sampling $\theta^i \sim P$ and accepting the result only if $\theta^i \in R_{X^i, Z^i}$. [1]

Due to our choice of conjugate prior for P, we have

$$P \mid \theta^1, ..., \theta^n \sim DP(\Theta, \alpha + \sum_{i=1}^{n} \delta_{\theta^i}) \tag{8}$$

We can sample P from this distribution using a stick breaking process.

Together (7) and (8) define a Gibbs sampler that be used to sample P and $(\theta^1, ..., \theta^n)$ from their posterior distributions. Although our intention is to learn the posterior of P, we cannot store a sample of distributions for P. Instead, it will prove sufficient to take a sample for $(\theta^1, ..., \theta^n)$.

Amalgamating the two stages of the Gibbs Sampler, successive instances of $(\theta^1, ..., \theta^n)$ are computed by the following procedure.

For $i = 1, ..., n$ {

Suppose we have already obtained stick breaking elements $(p_1, \tilde{\theta}_1), ..., (p_k, \tilde{\theta}_k)$.

1. Sample $U \sim U(0, 1)$.
2. If $U > \sum_{r=1}^{k} p_r$, generate $(p_{k+1}, \tilde{\theta}_1), ..., (p_{k'}, \tilde{\theta}_{k'})$ until k' is minimal such that $U \leq \sum_{r=1}^{k'} p_r$ as follows:

   For $j = k + 1, k + 2, ...$ {

   i. Sample $\beta_j \sim Beta(1, \alpha(\mathtt{R}) + n)$. Set $p_j = \beta_j \prod_{r=1}^{j-1}(1 - p_{j'})$.

[1] If R_{X^i, Z^i} is small, this rejection sampling can be very slow. That issue is addressed in Section 6.

ii. Sample $\tilde{\theta}_j$ from the distribution

$$\frac{1}{(\alpha(\mathtt{R})+n)}\left(\alpha+\sum_{i=1}^{n}\delta_{\theta^i}\right)$$

i.e. with probability $n/(\alpha(\mathtt{R})+n)$ we sample $\tilde{\theta}_j$ uniformly at random from $\theta^1,...,\theta^n$. Else we sample $\tilde{\theta}_j$ from the distribution $\alpha/\alpha(\mathtt{R})$.

}

3. Set $\tilde{\theta}=\tilde{\theta_{k'}}$. If $\tilde{\theta}\in R_{X_i,Z_i}$, set $\theta^{i^*}=\tilde{\theta}$. Else return to 1.

}

We allow for some burn-in, then construct a sample for $(\theta^1,...,\theta^n)$ by selecting distantly spaced instances from this sequence.

5 Prediction

Now for prediction we suppose that we have a new agent on whom we have historic data X', Z'. We are interested in the probability they will accept a suggested prospect X^*

$$p^* = p(Z^*=1|X^*,X',Z') \tag{9}$$

The randomness modelled by p^* here is whether the given agent is the type person who would accept this prospect. The decision Z^* is determined by the agent's parameter θ' and we consider θ' to be drawn from P, conditional on the information (X,Z). In particular, p^* is fixed given P.

There is a second layer of randomness relating to our uncertainty in P, encapsulated by our posterior for P given D. We will let p^* inherit a posterior distribution given D also. Treating this randomness seperately allows us to assess our uncertainty due to imperfect training of the algorithm, as distinct from the fundamental uncertainty resulting from population heterogeneity.

$Z^*=1$ if and only if $\theta'\in R_{X^*}$.

From (6), we have

$$\begin{aligned} p^* &= p(\theta'\in R_{X^*}|\{\theta'\in R_{X',Z'}\}) \\ &= P|_{R_{X',Z'}}(R_{X^*}) \end{aligned} \tag{10}$$

Then, by (8),

$$p^*\,|\,\theta^1,...,\theta^n \sim \mathrm{Beta}\,(\,K(R_{X^*}\cap R_{X',Z'}),\,K(R^c_{X^*}\cap R_{X',Z'})\,) \tag{11}$$

where $K(A) = \alpha(A) + \#\{i : \theta^i \in A\}$. Note that K is staight-forward to evaluate for these sets. It simply requires us to check whether each θ^i is in the sets R_{X^*} and $R_{X',Z'}$.

This shows how the posterior distribution of p^* may be inferred from our sample for $(\theta^1, ..., \theta^n)$. For each $(\theta^1, ..., \theta^n)$, we sample one instance of p^* from (11), and so obtain a sample for p^*. In fact, values for the mean and variance of $p^* \mid D$ are likely to be sufficient, as an estimate of the desired probability and a measure of our training uncertainty. These may be estimated directly from the sample for $(\theta^1, ..., \theta^n)$ using

$$\begin{aligned} E(p^* \mid D) &= E_{\theta^1,...,\theta^n} \left\{ E(p^* \mid \theta^1, ..., \theta^n) \mid D \right\} \\ &= E_{\theta^1,...,\theta^n} \left\{ \frac{K(R_{X^*} \cap R_{X',Z'})}{K(R_{X',Z'})} \right\} \end{aligned} \tag{12}$$

$$\begin{aligned} Var(p^* \mid D) &= Var_{\theta^1,...,\theta^n} \left\{ E(p^* \mid \theta^1, ..., \theta^n) \mid D \right\} \\ &\quad + E_{\theta^1,...,\theta^n} \left\{ Var(p^* \mid \theta^1, ..., \theta^n) \mid D \right\} \\ &= Var_{\theta^1,...,\theta^n} \left\{ \frac{K(R_{X^*} \cap R_{X',Z'})}{K(R_{X',Z'})} \right\} \\ &\quad , + E_{\theta^1,...,\theta^n} \left\{ \frac{K(R_{X^*} \cap R_{X',Z'}) K(R^c_{X^*} \cap R_{X',Z'})}{K(R_{X',Z'})^2 (K(R_{X',Z'}) + 1)} \right\} \end{aligned} \tag{13}$$

6 Robusting the algorithm

As with any model, some agents will be poorly described by Prospect Theory. Indeed, PT does not constitute a perscription for optimal decision making under some criteria. Rather, it is a psychological assertion about how typical agents perceive their situation and it is inevitable that some agents will behave differently.

This section focuses on refining the algorithm to handle agents who deviate from PT. Such agents present a number of problems for the algorithm. Firstly, consider an agent in the training set whose responses Z^i to prospects X^i do not fit with PT. In that case, the region R_{X^i,Z^i} is likely to be very small or even empty. Using the likelihood for θ given in (5), we note that the likelihood for P is

$$L(P) \propto \prod_{i=1}^{n} P(R_{X^i,Z^i}) \tag{14}$$

If R_{X^i,Z^i} is small, the posterior for P will be biased towards distributions that assign significant mass to this erronenous region. If the set is empty, this likelihood is everywhere 0, so the posterior is not even defined.

Further, while training the algorithm, the Gibbs sampler must sample values for θ^i until we obtain $\theta^i \in R_{X^i,Z^i}$. If R_{X^i,Z^i} is small, this will take a very long time. Indeed, if it is empty the algorithm will not terminate at all.

On the other hand, during prediction suppose that the agent in question does not follow PT closely. Then of course we cannot predict their actions well, so our distribution for p^* ought to be centred near 1/2 with a large variance. In fact, since the agent's previous responses Z' to prosepcts X' will not fit with PT, we are likely to have a region $R_{X',Z'}$ that is very small. Consequently, we expect one of the following two situations to hold approximately:

1. $$R_{X',Z'} \subset R_{X^*}$$

2. $$R_{X',Z'} \subset R^c_{X^*}$$

It follows from (11) that the posterior for p^* will be clustered closely around either 0 or 1.

A number of recent studies have addressed the issue that different agents are better described by different decision models using mixture models [Bruhin et al. (2010), Harrison and Rutström (2009), Conte et al. (2010)]. For instance, Bruhin et al. uses a mixture model that classifies agents as either following PT or EUT. They find most agents have a very high posterior probability of belonging to one of the two classes, with about 80% of agents following PT.

In principle, our predictor could be combined with other predictors based on alterative decision models in this way. In fact, we will presume that in the context under consideration almost all agents follow PT reasonably well, and let only some small proportion ϵ deviate from it. We will not specify an alternative model for those agents. Rather, we assert that they follow some reference model, which we do not attempt to learn, that predicts each decision correctly with a given probability $\phi \in [1/2, 1]$[2]. We will consider ϕ to be the same for all training agents. However, it will be convenient to use a different value of ϕ for prediction.

To model this, we endow each agent with a binary variable Y that indicates whether the agent deviates from PT.

$$Y \sim Bernoulli(\epsilon) \tag{15}$$

independently across agents.

When we make prediction, it will be important that PT is compared only against an imperfect reference model. For, even if the agent in question follows PT with

[2]We suppose $\phi \geq 1/2$ so that the reference model is at least as accurate as guessing at random.

some θ exactly, the population heterogeneity means that PT will not be able to predict their decisions perfectly until we have learnt θ. Meanwhile, when compared against a perfect model, any decision history (X', Z') would appear as evidence that the agent deviates from PT. Once m' is sufficiently large, we are sure to conclude that $Y = 1$ for this agent.

For training, however, the details of the reference model are less important and we are free to use $\phi = 1$. In this case, our choice of ϕ essentially controls the extent to which our inference on P is weighted towards training agents for whom we have more information. Reducing ϕ moves the posterior for P away from those agents where m is large.

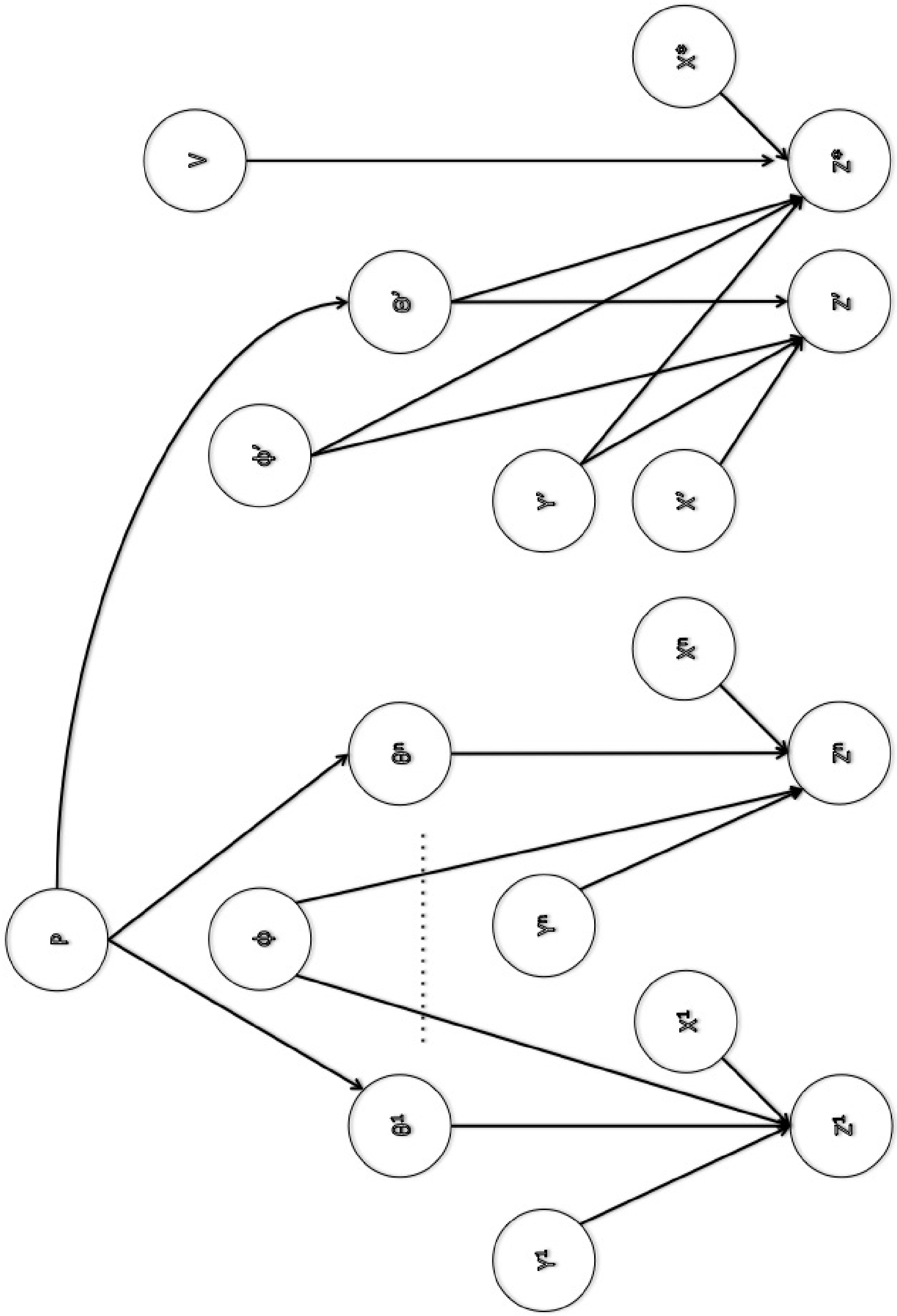
For observed data (X, Z), we now have the conditional sampling distributions

$$p(Z \mid X, \{Y = 0\}, \theta) = 1(\theta \in R_{X,Z}) \tag{16}$$

$$p(Z \mid X, \{Y = 1\}, \phi) = \phi^m \tag{17}$$

After marginalising out Y, this gives the robust likelihood for θ (cf (5))

$$L(\theta) = (1 - \epsilon)1(\theta \in R_{X,Z}) + \phi^m \epsilon \tag{18}$$



6.1 Training

Using (18), we obtain the robust likelihood for P (cf (14))

$$L(P) \propto \prod_{i=1}^{n} \left\{ P(R_{X^i,Z^i}) + \frac{\phi^m \epsilon}{1-\epsilon} \right\} \tag{19}$$

Now $P(R_{X^i,Z^i})$ can be small for some agent without the likelihood approaching zero. This shows that no single agent can bias the posterior for P too heavily.

The algorithm is trained using the same Gibbs sampler as before. The difference is that during the rejection sampling we must consider the possibility that $Y = 1$ for the given agent. Step 3 is replaced by

3' Set $\tilde{\theta} = \tilde{\theta_{k'}}$. If $\tilde{\theta} \in R_{X_i,Z_i}$, set $\theta^{i^*} = \tilde{\theta}$.

4' If $\tilde{\theta} \notin R_{X_i,Z_i}$, sample $T \sim Bernoulli(\phi^{m^i}\epsilon)$[3]. If $T = 1$, set $\theta^{i^*} = \tilde{\theta}$. Else return to 1.

Note that the rejection sampling now requires at most $O(1/\phi^{m^i}\epsilon)$ attempts, so the run-time is bounded even if R_{X^i,Z^i} is small.

6.2 Prediction

Equation (17) describes how the reference model fits observed decisions. However, since we do not learn the reference model, given a new prospect X^* we do not know it should predict the response Z^*. We define a random variable V for this predicted response and model it as

$$V \sim Bernoulli(1/2) \tag{20}$$

independent of X', Z', D.

Since our inability to predict V reflects a limitation of the algorithm, we want to include this additional source of randomness as part of our posterior uncertainty over p^*. Namely, we now consider p^* to be fixed given P, V. p^* then inherits a posterior distribution from our posterior for P and the variability in V.

To evaluate this distribution, we first compute

$$p(Y' = 1 \mid X', Z', \phi', \theta') = \begin{cases} \frac{\phi^{m'}\epsilon}{1-(1-\phi^{m'})\epsilon} & \theta' \in R_{X',Z'} \\ 1 & \theta' \notin R_{X',Z'} \end{cases} \tag{21}$$

$$\therefore p(Y' = 1 \mid X', Z', \phi') = 1 - \left\{ \frac{1-\epsilon}{1-(1-\phi^{m'})\epsilon} \right\} P(R_{X',Z'}) \tag{22}$$

Then we have

[3] $\phi^{m^i}\epsilon$ is the probability that $Y = 1$ and the observed response Z is given.

$$
\begin{aligned}
p^* &= p(Y'=0\,|\,X',Z',\phi')\,p(\theta'\in R_{X^*}\,|\,\{Y'=0\},\{\theta'\in R_{X',Z'}\}) \\
&\quad + p(Y'=1\,|\,X',Z',\phi')\,V \\
&= \left\{\frac{1-\epsilon}{1-(1-\phi^{m'})\epsilon}\right\} P(R_{X',Z'})\left\{P|_{R_{X',Z'}}(R_{X^*})-V\right\}+V
\end{aligned}
\tag{23}
$$

By (8),

$$
P(R_{X',Z'})\,|\,\theta^1,...,\theta^n \sim \mathrm{Beta}\left(\,K(R_{X',Z'}),\;K(R^c_{X',Z'})\,\right) \tag{24}
$$

$$
P|_{R_{X',Z'}}(R_{X^*})\,|\,\theta^1,...,\theta^n \sim \mathrm{Beta}\left(\,K(R_{X^*}\cap R_{X',Z'}),\;K(R^c_{X^*}\cap R_{X',Z'})\,\right) \tag{25}
$$

hold independently.

By sampling from (20), (24) and (25) for each $(\theta^1,...,\theta^n)$ in our training sample, we can generate a complete sample for p^* from its posterior distribution. Alternatively we may compute the posterior mean and variance of p^* as before.

$$
\begin{aligned}
E(p^*\,|\,D) &= E_{\theta^1,...,\theta^n}\left\{E(p^*\,|\,\theta^1,...,\theta^n)\,|\,D\right\} \\
&= E_{\theta^1,...,\theta^n}\left\{\left(\frac{1-\epsilon}{1-(1-\phi'^{m'})\epsilon}\right)\left(\frac{K(R_{X',Z'})}{K(\mathtt{R})}\right)\right. \\
&\qquad \left.\left[\left(\frac{K(R_{X^*}\cap R_{X',Z'})}{K(R_{X',Z'})}\right)-\frac{1}{2}\right]+\frac{1}{2}\right\} \\
&= \frac{1}{2}+\left\{\frac{1-\epsilon}{[\alpha(\mathtt{R})+n]\,[1-(1-\phi'^{m'})\epsilon]}\right\} \\
&\qquad E_{\theta^1,...,\theta^n}\left\{K(R_{X^*}\cap R_{X',Z'})-\frac{1}{2}K(R_{X',Z'})\right\}
\end{aligned}
\tag{26}
$$

A similar (but very messy) formula can be obtained for the variance.

It is worth noting that if $R_{X',Z'}$ is very small, then $p^*\approx V$. Rather than concluding that the agent's decision may be predicted with great confidence, we find that p^* is centred around 1/2 with maximum possible variance 1/4.

Finally we address the issue of choosing ϕ for prediction. Supposing that the agent is well-described by PT, the region $R_{X',Z'}$ will shrink with increasing m' as we learn their value for θ. From (22) we see that (conditional on P) the population heterogeneity ultimately leads us to conclude that the agent deviates from PT if $P(R_{X',Z'})=o\left(\phi'^{m'}\right)$ as $m'\to\infty$.

Consequently, we would like to choose ϕ' so that $P(R_{X',Z'})\sim\phi'^{m'}$ for typical distributions P under the posterior. This could be achieved, perhaps, by examining how $P(R_{X,Z})$ decays for agents in the training set. Alternatively, to be safe, we could just use $\phi' = 1/2$.

7 References

Abdellaoui, M., Bleichrodt, H., Paraschiv, C. (2007), *Loss Aversion under Prospect Theory: A Parameter-Free Measurement*, Management Science 50(10), 1659-1674.

Bali, T., Cakici N., Whitelaw, R. (2011), *Maxing Out: Stocks as Lotteries and the Cross-section of Expected Returns*, Journal of Financial Economics 99(2): 427-446.

Barberis, N., Huang, M. (2008), *Stocks as Lotteries: The Implications of Probability Weighting for Security Prices*, American Economic Review 98(5): 2066-2100.

Barberis, N. (2012), *A Model of Casino Gambling*, Management Science 58(1): 35-51.

Barberis, N. (2013), *Thirty Years of Prospect Theory in Economics: A Review and Assessment*, Journal of Economic Perspectives, forthcoming.

Barseghyan, L., Molinari, F., O'Donoghue, T., Teitelbaum, J. (2011), *The Nature of Risk Preferences: Evidence from Insurance Choices*, SSRN Working Paper 1646520.

Booij, A., van Praag, B., van de Kuilen, G. (2009), *Parametric Analysis of Prospect Theory's Functionals for the General Population*, IZA DP No. 4117.

Boyer, B., Mitton, T., Vorkink, K. (2010), *Expected Idiosyncratic Skewness*, Review of Financial Studies 23(1): 169-202.

Bruhin, A., Fehr-Duda, H., Epper, T. (2010), *Risk and Rationality: Uncovering Heterogeneity in Probability Distortion*, Econometrica, 78: 13751412.

Camerer, C., Babcock, L., Loewenstein, G., Thaler, R. (1997), *Labor Supply of New York City Cabdrivers: One Day at a Time* ,Quarterly Journal of Economics 112(2), 407-441.

Chajewska, U., Koller, D. (2000), *Utilities as random variables: Density estimation and structure discovery*, Proc. UAI, 63-71.

Chajewska, U., Koller, D. (2001), *Learning an agents utility function by observing behavior*, Proc. UAI, 35-42.

Conte, A., Hey J. D., Moffatt, P. G. (2010), *Mixture Models of Choice Under Risk*, Journal of Econometrics, 1378-1409.

Gelfand, A. E., Kottas, A. (2002), *A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models*, Journal of Computational and Graphical Statistics, 11:289-305.

Harrison, G.W., Rutström, E.E. (2009), *Representative Agents in Lottery Choice Experiments: One Wedding and a Decent Funeral*, Experimental Economics, 12, 133158.

Heidhues, P., Koszegi, B. (2012) *Regular Prices and Sales*, unpublished paper.

Hey, J. D., C. Orme (1994), *Investigating Generalizations of Expected Utility Theory Using Experimental Data*, Econometrica, 62, 12911326.

Koszegi, B., Rabin, M. (2006), *A Model of Reference-Dependent Preferences*, Quarterly Journal of Economics 121(4): 1133-1165.

MacEachern, S. N. (1998), *Computational methods for mixture of Dirichlet process models*, Practical Nonparametric and Semiparametric Bayesian Statistics, 23-44. Springer.

Neal, R. M. (2000), *Markov chain sampling methods for Dirichlet process mixture models*, Journal of Computational and Graphical Statistics, 9:249-265.

Sethuraman, J. (1994), *A constructive definition of Dirichlet priors*, Statistica Sinica 4: 639-650.

Snowberg, E., Wolfers, J. (2010), *Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions?*, Journal of Political Economy 118(4): 723-746.

Starmer, C., (2000), *Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk*, Journal of Economic Literature 38, 332382

Teh, Y. W., Jordan, M. I., Beal, M. J., Blei, D. M. (2004), *Hierarchical Dirichlet processes*, Technical Report 653, UC Berkeley.

Tversky, A., Kahneman, D. (1979), *Prospect Theory: An Analysis of Decision under Risk*, Econometrica, 47(2), pp. 263-291.

Tversky, A., Kahneman, D. (1992), *Advances in Prospect Theory: Cumulative Representation of Uncertainty*, Journal of Risk and Uncertainty 5, 297323.

Wakker, P.P. (1994), *Separating Marginal Utility and Probabilistic Risk Aversion*, Theory and Decision 36, 144.

Wakker, P.P. (2008) *Explaining the Characteristics of the Power (CRRA) Utility Function*, Health Economics, 17, 1329-1344.