Expert Information and Nonparametric Bayesian Inference of Rare Events

Hwan-sik Choi
Purdue University

hwansik@purdue.edu

Nov 2012
Summary

Objective: Find a good method for inference on rare events such as

- default probability or number of defaults in a high grade portfolio,
- value at risk, chance of extreme losses,
- sovereign default,
- disasters and other catastrophic events (a.k.a. tail-risk events).

We develop a nonparametric Bayesian framework for inference of rare events.

- We use the Dirichlet process mixture (DPM) model.
- Expert information is combined with an econometrician’s DP prior.
- We will discuss possible extensions to semiparametric models.
Motivation: Inference of Rare Events is Difficult

Problem: Lack of historical data information.

1. For a parametric model, checking model adequacy is difficult.  
   *Is the tail part of our model distribution correctly specified?*

2. For a Bayesian, non-informative or objective prior is not satisfactory.  
   *Data have little information on the tail.*

We argue

- nonparametric models are appealing because of Problem 1, and
- non-data information is appealing for Problem 2.
Use of Expert Information

Kiefer (2009, 2010)

- uses expert information for default estimation, and
- argues that the Bayesian approach is a natural and coherent way of combining multiple sources of information for defaults.

We develop a nonparametric model to handle the concerns of misspecification.

- How to elicit expert information when the dimension of the model is infinite?
  Experts can talk about some aspects of rare events only.
- How to combine expert information with an econometrician’s prior?
  Get the least informative (maximum entropy) prior that complies with expert knowledge.
Sampling distribution of $y$:

- Infinite mixture of a kernel function $K(y|\xi)$,

$$y|G \sim \int K(y|\xi) G(d\xi).$$

- Mixing measure $G$ is an infinite dimensional parameter.
- We need a prior distribution over the space $\mathcal{G}$ of all $G$.
- Dirichlet process has become popular as a prior on $\mathcal{G}$. 

Dirichlet Process Mixture Model
Dirichlet Process

Dirichlet process \( \mathcal{P} = \text{DP}(\alpha G_0) \) is a random measure:

- Base measure \( G_0 \in \mathcal{G} \) is the expected value, \( G_0 = \int G \mathcal{P}(dG) \).
- Concentration parameter \( \alpha > 0 \) is precision around \( G_0 \).
  
  If \( \alpha \to \infty \), \( G \overset{d}{\to} G_0 \).
- Its finite dimensional distribution is the Dirichlet distribution.
- A draw \( G \sim \text{DP}(\alpha G_0) \) is almost surely discrete.

Elicitation of \( \alpha \) and \( G_0 \):

- they are too difficult for an expert to think about, and
- their relationship with rare events may not be obvious.
How to Elicit Expert information?

Experts are interested in a vector $\varphi$ of some functionals of a sampling distribution $F$:

$$\theta = \varphi(F(y|G)).$$

- For example, $\theta = P\{y < 0|G\}$ is the probability of default if $y$ is an equity value.

- Elicitation of expert information: Ask ...
  - What is the level of $\theta$ above and below which are equally likely (i.e. median)?
  - What do you think is the probability $P\{\theta < 1\%\}$?
  - Quartiles? Have max or min?

- Avoid asking moments of $\theta$ since they are not easy or natural to think about.
Assumption: Expert information is given by moment conditions

\[ \mathbb{E} g(\theta) = 0. \]

Consider the space of priors

\[ \mathcal{Q} = \{ Q : \mathbb{E}^Q g(\theta) = 0 \} \]

that comply with expert knowledge.
Merging Expert Information

Find the prior $Q^* \in \mathcal{Q}$ closest to the DP prior $\mathcal{P}$ in Kullback-Leibler information criterion by

$$\min_{Q \in \mathcal{Q}} \text{KLIC}(Q \| \mathcal{P}).$$

Amari (1982) calls $\text{KLIC}(Q \| \mathcal{P})$ as $-1$-divergence from $\mathcal{P}$ to $Q$.

$Q^*$ is obtained by $-1$-projection of $\mathcal{P}$ onto the space $\mathcal{Q}$. 
Least Informative Prior with Expert Information

\( Q^* \) is given by Gibbs canonical density \( \pi^* \):

\[
\pi^* = \frac{dQ^*}{dP} = \frac{\exp(\lambda_* g(\theta))}{\mathbb{E}^P \exp(\lambda_* g(\theta))},
\]

where \( \lambda_* \) solves

\[
\min_{\lambda} \mathbb{E}^P \exp(\lambda' g(\theta))
\]

and

\[
D(Q^* \| P) = -\log(\mathbb{E}^P \exp(\lambda_*' g(\theta))).
\]

We call our approach ETDP (exponentially tilted DP).
Suppose $y \sim \mathcal{G}$ for simplicity. Consider moment conditions

$$\mathbf{E}^{\mathcal{G}} m(y, \theta) = 0,$$

where $\theta \in \Theta$. Consider the space

$$\mathcal{G}_\theta = \{ \mathcal{G} \in \mathcal{G} | \mathbf{E}^{\mathcal{G}} m(y, \theta) = 0 \}$$

and denote $\mathcal{G} = \bigcup_{\theta \in \Theta} \mathcal{G}_\theta$.

For each $\mathcal{G} \in \mathcal{G}$, define $\theta = \{ \theta | \mathcal{G} \in \mathcal{G}_\theta \}$, and $\theta = \emptyset$ for $\mathcal{G} \in (\mathcal{G} - \mathcal{G})$.

Assume that $\theta$ is fully identified, i.e. $\mathcal{G}_\theta \cap \mathcal{G}_{\theta'} = \emptyset$ for $\theta \neq \theta'$. 
Semiparametric Bayesian Models

ET-ETD-DP (Exponential tilting of exponentially tilted draws of DP).

- $-1$-projection $G^*_{\theta}$ of $G \sim \mathcal{P}$ onto $\bar{G}$ to get the degenerated prior $\mathcal{P}^*$
- $-1$-projection $Q^*$ of $\mathcal{P}^*$ onto $Q$ to merge expert information.

Alternatively, we can use Kitamura and Otsu (2011):

- Define a marginal prior $p(\theta)$ on $\theta$ and
- Given $\theta$, use $-1$-projection of $G \sim \text{DP}(\alpha G_0)$ onto $G_\theta$ for $p(G|\theta)$.
- ET-CETD-DP (Exponential tilting of conditionally exponentially tilted draws of DP)?

If $\bar{G} = G$, then we can directly use ETDP.
Geometry of Using Expert Information

\[ DP \text{ prior } + \text{Moment info. } + \text{Expert info. } + \text{Data} = \text{Posterior} \]
Estimation of Gibbs Density

Given $\mathcal{P} = \text{DP}(\alpha G_0)$,

- simulate $\theta_1, \ldots, \theta_M$ for a large $M$ by drawing $G \sim \text{DP}(\alpha G_0)$,
- solve the maximum entropy problem for $\{\pi_m\}_{m=1}^M$

$$\max_{(\pi_1, \ldots, \pi_M)} \sum_{m=1}^M \pi_m \log(1/\pi_m)$$

such that

$$\sum_{m=1}^M \pi_m = 1, \quad \sum_{m=1}^M \pi_m g(\theta_m) = 0.$$
Exponential Tilting of Dirichlet Prior

The solution \( \{ \hat{\pi}_m \} \) is given by the exponential tilting estimator of Kitamura and Stutzer (1997)

\[
\hat{\pi}_m = \frac{\exp \left( \hat{\lambda}_M' g(\theta_m) \right)}{\sum_{m=1}^{M} \exp \left( \hat{\lambda}_M' g(\theta_m) \right)},
\]

where \( \hat{\lambda}_M \) is the Lagrange multiplier of the constraints, and

\[
\hat{\lambda}_M = \arg\min_{\lambda \in \Lambda} M^{-1} \sum_{m=1}^{M} \exp(\lambda' g(\theta_m))
\]

\( \xrightarrow{p} \lambda_* \) as \( M \to \infty \).
Posterior Simulation

Posterior distribution:

\[ \Psi^*(dG|y) \propto \prod_{i=1}^{n} f(y_i|G)\pi^*(\theta)P(dG) = \prod_{i=1}^{n} f(y_i|G)Q^*(dG). \]

To sample from the target distribution \( \Psi^* \), we apply the independence chain Metropolis-Hastings (M-H) algorithm.

- It is easy to sample from the posterior distribution \( \Psi(dG|y) \) with the original DP prior.
- So we use \( \Psi(dG|y) \) as the proposal distribution.
- Use the blocked Gibbs algorithm of Ishwaran and James (2001) for \( \Psi(dG|y) \).
Metropolis-Hastings Step

Using the relationship

\[ \frac{\Psi^*(dG|y)}{\Psi(dG|y)} = \pi^* , \]

we get the M-H acceptance probability of \( t \)-th MCMC sample of \( G(t) \)

\[ A(t) = \frac{\pi^*(\theta(t))}{\pi^*(\theta(t-1))} \approx \frac{\exp\left(\hat{\lambda}_M' g(\theta(t))\right)}{\exp\left(\hat{\lambda}_M' g(\theta(t-1))\right)} \]

\( G(t) \) is accepted with probability \( \min\{1, A(t)\} \),
or \( G(t) = G(t-1) \) if rejected.
Demonstration with Simulated Data

What is the probability of default over one year horizon?

Let’s assume that the true distribution of the future equity value \( y \) is mixture normal,

\[
y \sim 0.5 \mathcal{N}(10, 5^2) + 0.5 \mathcal{N}(20, 5^2).
\]

We generate \( n = 20 \) i.i.d. samples from this.
Dirichlet Process Prior

We use the DPM model:

- with the normal distribution kernel $K(y|\xi) = \Phi(y|\mu, 1/\tau)$,
- mixing measure $G$ is a distribution on $(\mu, \tau)$, and
- use the normal-gamma distribution for $G_0$ on $(\mu, \tau)$.

Prior: $\text{DP}(\alpha G_0)$ with $\alpha = 10$, and for $G_0$, we use the conjugate normal-gamma distribution

$$G_0 \sim \text{NG}(\mu_0 = 15, n_0 = 1, \nu_0 = 6, \sigma_0^2 = 20).$$

We do inference on the probability $\theta$ of default $y < 0$,

$$\theta = \varphi(F(y|G)) = F(0|G).$$
Expert Information

Expert information:

- It is equally likely that the chance of the extreme loss is greater or smaller than 1%:
  \[ P\{\theta < 0.01\} = 50\%. \]

- Then given \( \theta \leq 1\% \), the level of \( \theta \) above and below which are equally likely is 0.5%:
  \[ P\{\theta < 0.005\} = 25\%. \]

Exponential tilting of the DP with the estimated Lagrange multiplier

\[ \hat{\lambda}_M' = (1.010, 1.285). \]
Posterior Distribution of $\theta$

Posterior Density of $\theta$

Posterior CDF of $\theta$
Predictive Distribution of $y$

**Predictive Density of $y$**

- DP
- Expert
- Data

**Predictive CDF of $y$**

- DP
- Expert
- Data
Posterior Distribution of the Number of Defaults

\[ H_k^n = \binom{n}{k} \theta^k (1 - \theta)^{n-k}. \]
Application to Portfolio Returns

What is the probability of an extreme loss of a portfolio over the next year?

Build a homogeneous portfolio with the following characteristic as of 6/30/11.

- Market capitalization: $10-30MM
- Price per share / book value per share: 1-4
- Market beta: 0.8-1

Data from companies in S&P 500 index: We find 16 companies with the above characteristics.
Data: 6/30/11 - 6/29/12

Annual returns of 16 companies in a homogeneous portfolio.

<table>
<thead>
<tr>
<th>Company</th>
<th>Return (%)</th>
<th>Company</th>
<th>Return (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aetna Inc</td>
<td>−12.1</td>
<td>Raytheon Co</td>
<td>13.5</td>
</tr>
<tr>
<td>Archer-Daniels-Midland Co</td>
<td>−2.1</td>
<td>Republic Services Inc</td>
<td>−14.2</td>
</tr>
<tr>
<td>Chubb Corp/The</td>
<td>16.3</td>
<td>Staples Inc</td>
<td>−17.4</td>
</tr>
<tr>
<td>Humana Inc</td>
<td>−3.8</td>
<td>Stryker Corp</td>
<td>−6.1</td>
</tr>
<tr>
<td>Kohl’s Corp</td>
<td>−9.0</td>
<td>Symantec Corp</td>
<td>−25.9</td>
</tr>
<tr>
<td>Marsh &amp; McLennan Cos Inc</td>
<td>3.3</td>
<td>Thermo Fisher Scientific Inc</td>
<td>−19.4</td>
</tr>
<tr>
<td>Mylan Inc/PA</td>
<td>−13.4</td>
<td>WellPoint Inc</td>
<td>−19.0</td>
</tr>
<tr>
<td>Northrop Grumman Corp</td>
<td>−8.0</td>
<td>Zimmer Holdings Inc</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Max = 16.3%, Min = −25.9%, Mean = −7.21%, Median = −8.53%
Standard deviation = 11.75%
Expert Information and DP prior

The rare event of interest is an extreme loss:

\[
\theta = P\{y < -30|G\}.
\]

- Expert information \( P\{\theta < 0.02\} = 50\% \) and \( P\{\theta < 0.01\} = 25\% \).
- Prior: \( \text{DP}(\alpha G_0) \) with \( \alpha = 10 \), and
  \( G_0 \sim \text{NG}(\mu_0 = 0, \, n_0 = 1, \, \nu_0 = 6, \, \sigma_0^2 = 100) \).

Exponential tilting with the estimated Lagrange multiplier

\[
\hat{\lambda}_M' = (0.535, \, 1.315).
\]
Posterior Distribution of Probability of Extreme Loss

Posterior Density of $\theta$

Posterior CDF of $\theta$

Expert

DP
Predictive Distribution of Returns

Predictive Density of $y$

Predictive CDF of $y$
Posterior Distribution of the Number of Extreme Losses

Posterior Density of $H_{16}$

Posterior CDF of $H_{16}$
Conclusion

Inference of rare events:

- Misspecification? ⇒ Nonparametric model.
- Combining expert information? ⇒ Bayesian approach.

Inference of rare events with some new developments in Bayesian nonparametrics such as

- New MCMC methods: retrospective MCMC, slice sampling.
- Application to semiparametric models: estimating functions, GMM.
- Conditional models with covariates: DP that depends on covariates, Bayesian density regression.