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.

Dirichlet Process Mixture Model

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

Dirichlet process $\mathcal{P} = \mathsf{DP}(\alpha G_0)$ 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 $\alpha \to \infty$, $G \stackrel{d}{\to} G_0$.
- Its finite dimensional distribution is the Dirichlet distribution.
- A draw $G \sim \mathsf{DP}(\alpha G_0)$ is almost surely discrete.

Elicitation of α 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 φ of some functionals of a sampling distribution F:

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

- For example, $\theta = \mathbf{P}\{y < 0|G\}$ is the probability of default if y is an equity value.
- Elicitation of expert information: Ask ...
  - What is the level of θ above and below which are equally likely (i.e. median)?
  - What do you think is the probability $\mathbf{P}\{\theta < 1\%\}$?
  - Quartiles? Have max or min?
- Avoid asking moments of θ since they are not easy or natural to think about.

Priors in Expert's Mind

Assumption: Expert information is given by moment conditions

$$\mathbf{E}g(\theta) = 0.$$

Consider the space of priors

$$\mathbb{Q} = \{\mathcal{Q} : \mathbf{E}^{\mathcal{Q}} g(\theta) = 0\}$$

that comply with expert knowledge.

Merging Expert Information

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

$$\min_{\mathcal{Q} \in \mathbb{Q}} \mathsf{KLIC}(\mathcal{Q} \parallel \mathcal{P}).$$

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

$\mathcal{Q}^*$ is obtained by -1-projection of $\mathcal{P}$ onto the space $\mathbb{Q}$.

Least Informative Prior with Expert Information

$\mathcal{Q}^*$ is given by Gibbs canonical density π^*:

$$\pi^* = \frac{d\mathcal{Q}^*}{d\mathcal{P}} = \frac{\exp(\lambda_*' g(\theta))}{\mathbf{E}^{\mathcal{P}} \exp(\lambda_*' g(\theta))},$$

where λ_* solves

$$\min_{\lambda} \mathbf{E}^{\mathcal{P}} \exp(\lambda' g(\theta))$$

and

$$D(\mathcal{Q}^* \parallel \mathcal{P}) = -\log(\mathbf{E}^{\mathcal{P}} \exp(\lambda_*' g(\theta))).$$

We call our approach ETDP (exponentially tilted DP).

Semiparametric Models

Suppose $y \sim G$ for simplicity. Consider moment conditions

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

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

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

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

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

Assume that θ is fully identified, i.e. $\mathcal{G}_\theta \cap \mathcal{G}_{\theta'} = \varnothing$ for $\theta \neq \theta'$.

Semiparametric Bayesian Models

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

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

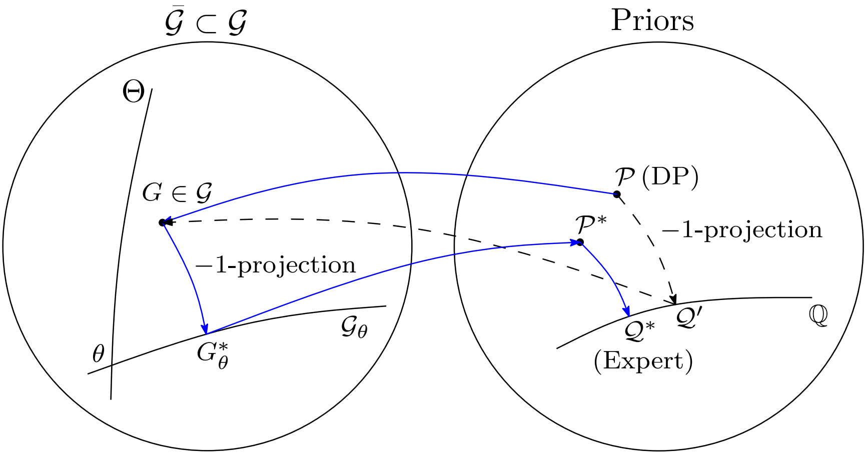
Alternatively, we can use Kitamura and Otsu (2011):

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

If $\bar{\mathcal{G}} = \mathcal{G}$, then we can directly use ETDP.

Geometry of Using Expert Information

DP prior + Moment info. + Expert info. + Data = Posterior



Estimation of Gibbs Density

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

- simulate $\theta_1, \ldots, \theta_M$ for a large M by drawing $G \sim \mathsf{DP}(\alpha G_0)$,
- 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

$$\sum_{m=1}^{M} \pi_m = 1, \qquad \sum_{m=1}^{M} \pi_m g(\theta_m) = 0.$$

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

$$\begin{aligned} \hat{\lambda}_M &= \underset{\lambda \in \Lambda}{\operatorname{argmin}}\, M^{-1} \sum_{m=1}^{M} \exp(\lambda' g(\theta_m)) \\ &\xrightarrow{p} \lambda_* \text{ as } M \to \infty. \end{aligned}$$

Posterior Simulation

Posterior distribution:

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

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

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

Metropolis-Hastings Step

Using the relationship

$$\frac{\Psi^*(dG|\mathbf{y})}{\Psi(dG|\mathbf{y})} = \pi^*,$$

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,

$$y \sim 0.5\mathbf{N}(10, 5^2) + 0.5\mathbf{N}(20, 5^2).$$

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 (μ, τ), and
- use the normal-gamma distribution for G_0 on (μ, τ).

Prior: DP(αG_0) with $\alpha = 10$, and for G_0, we use the conjugate normal-gamma distribution

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

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

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

Expert Information

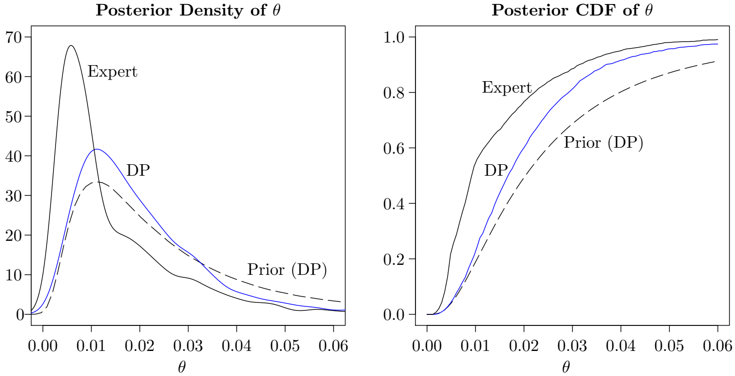
Expert information:

- It is equally likely that the chance of the extreme loss is greater or smaller than 1%:
$$\mathbf{P}\{\theta < 0.01\} = 50\%.$$
- Then given $\theta \leq 1\%$, the level of θ above and below which are equally likely is 0.5%:
$$\mathbf{P}\{\theta < 0.005\} = 25\%.$$

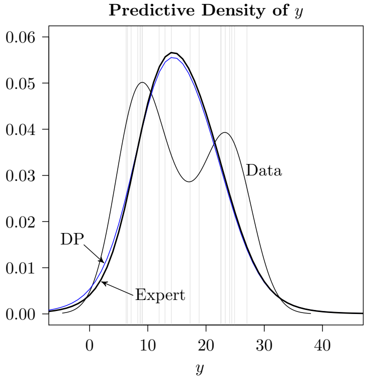
Exponential tilting of the DP with the estimated Lagrange multiplier

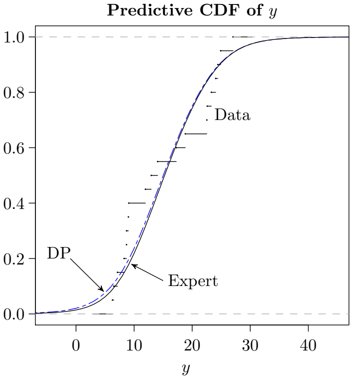
$$\hat{\lambda}'_M = (1.010,\ 1.285).$$

Posterior Distribution of θ



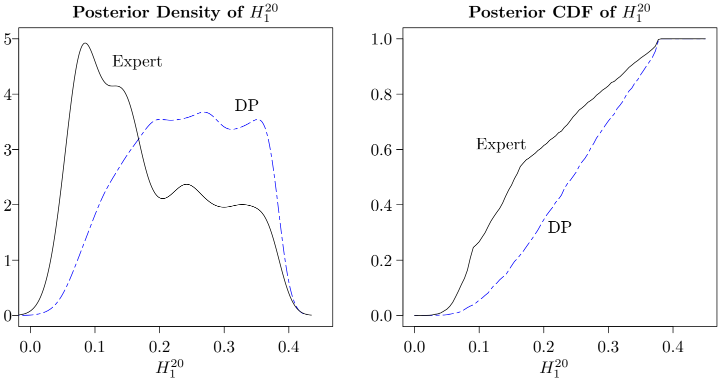
Predictive Distribution of y





Posterior Distribution of the Number of Defaults

$$H_k^n = \binom{n}{k}\theta^k(1-\theta)^k.$$



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.

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

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:

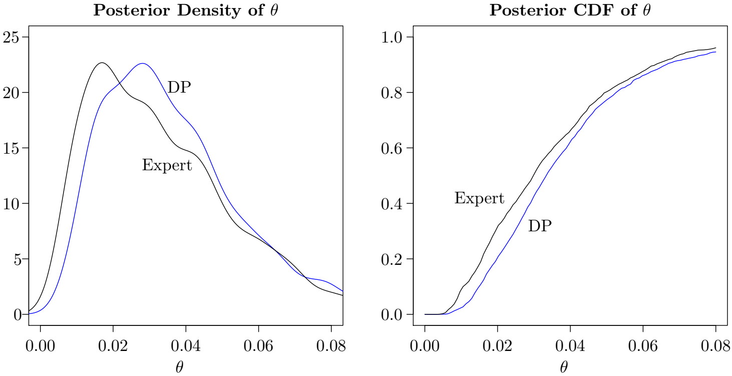
Probability that the annual return is -30% or worse
$$\theta = \mathbf{P}\{y < -30|G\}.$$

- Expert information $\mathbf{P}\{\theta < 0.02\} = 50\%$ and $\mathbf{P}\{\theta < 0.01\} = 25\%$.
- Prior: DP(αG_0) with $\alpha = 10$, and
  $G_0 \sim \mathbf{NG}(\mu_0 = 0,\ n_0 = 1,\ \nu_0 = 6,\ \sigma_0^2 = 100)$.

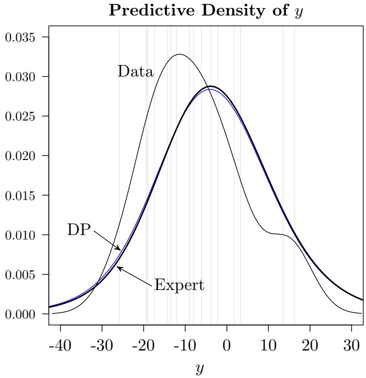
Exponential tilting with the estimated Lagrange multiplier

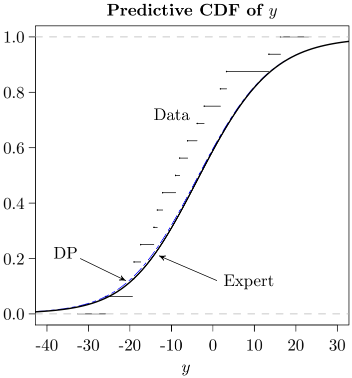
$$\hat{\lambda}'_M = (0.535,\ 1.315).$$

Posterior Distribution of Probability of Extreme Loss

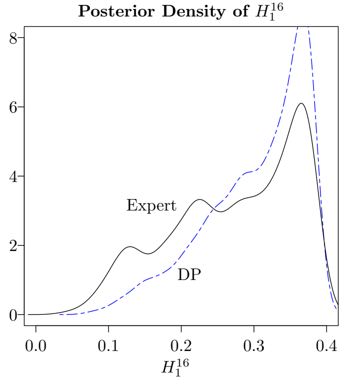


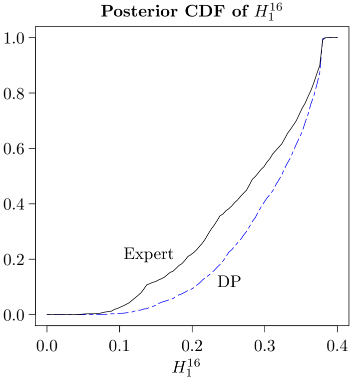
Predictive Distribution of Returns





Posterior Distribution of the Number of Extreme Losses





Conclusion

Inference of rare events:

- Misspecification? $\Rightarrow$ Nonparametric model.
- Lack of information? $\Rightarrow$ Expert information.
- Combining expert information? $\Rightarrow$ Bayesian approach.

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

- More general priors: Poisson-DP, stick-breaking processes.
- 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.