Approximate Inference in Industry: Two Applications at Amazon

Cédric Archambeau

cedrica@amazon.com

Workshop on Advances in Approximate Bayesian Inference,
Neural Information Processing Systems, 2017

Amazon.com, 2017

Today, machine learning is creating a paradigm shift.

"It is a golden age. Machine learning and AI is a horizontal enabling layer. It will empower and improve every business, every government organization, every philanthropy."

Jeff Bezos in Geekwire (May 6, 2017)

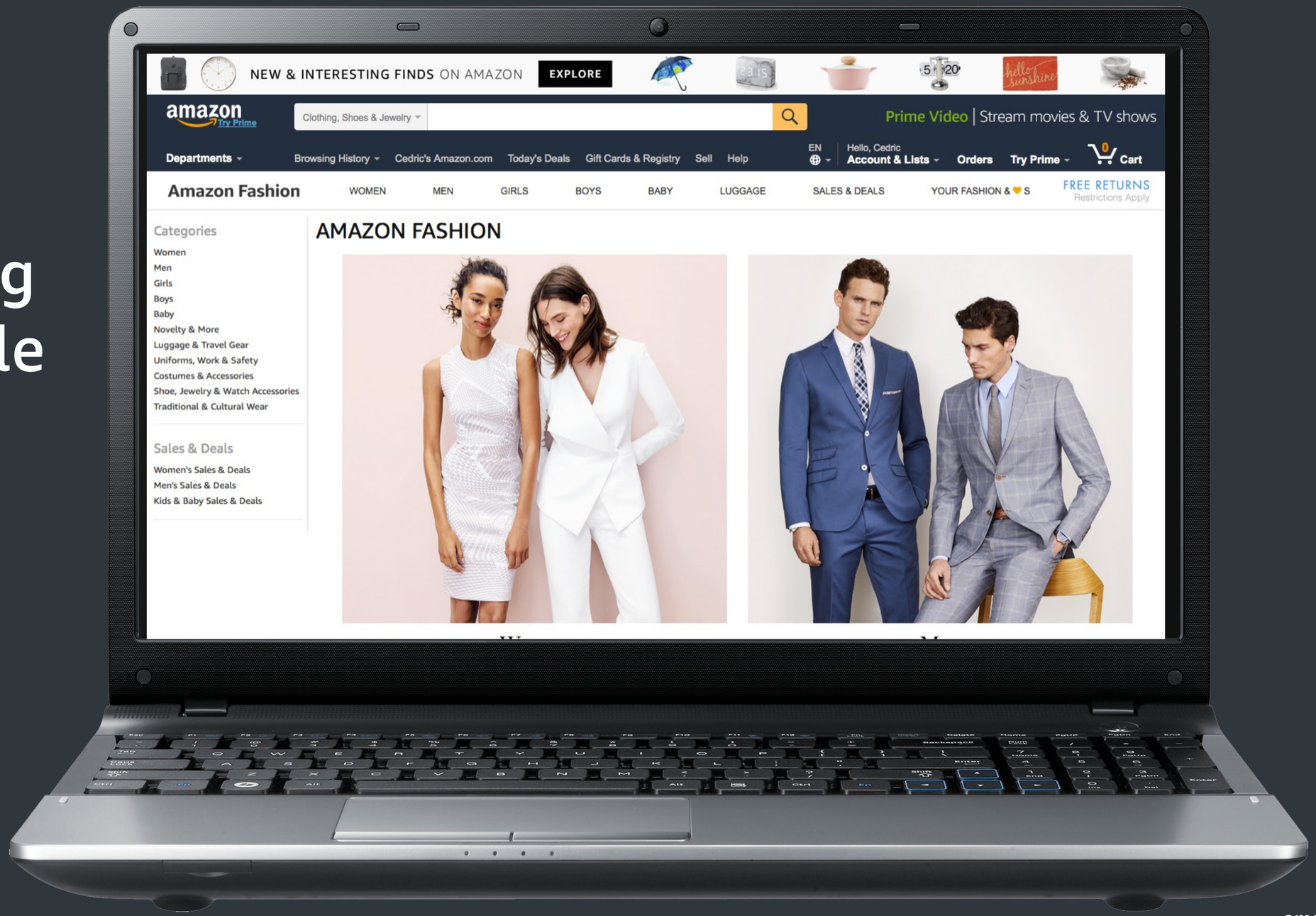
amazon

$$1 + 1 =$$

$$[(27/3)/3] - 1$$

amazon

Understanding Fashion & Style

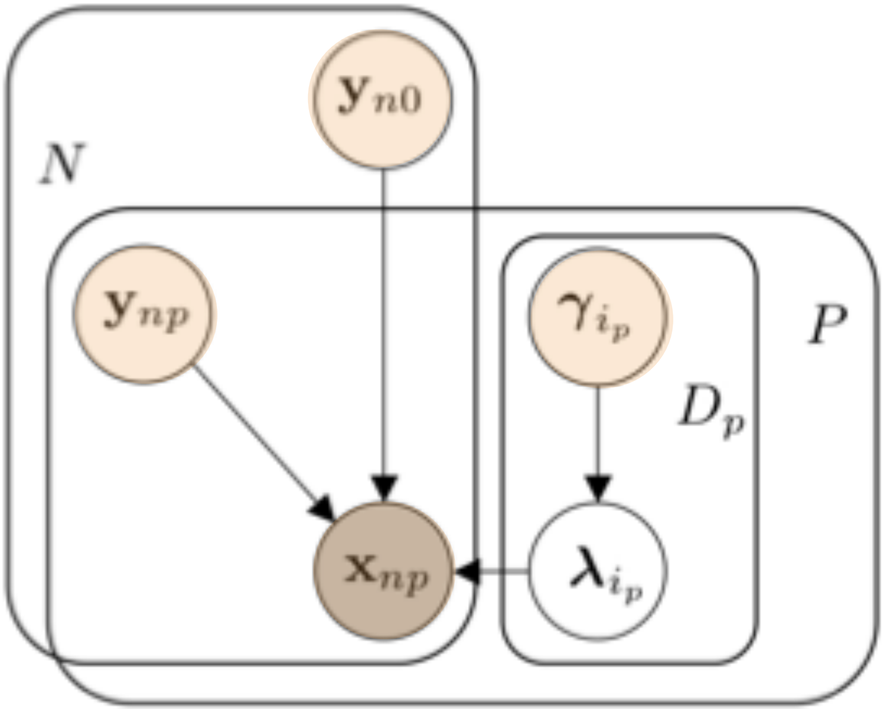


B00GIZOIDQ

null

Hair Drawing #4

amazon



(Archambeau and Bach, NIPS 2008)

Shallow Latent Variable Model

amazon

Variational Inference

- Using *Jensen's inequality*, we get for any distribution $q_{\mathbf{w}}(\mathbf{Z}, \boldsymbol{\theta})$:

$$\begin{aligned}
\ln p(\mathbf{X}) &= \ln \iint p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) \, d\mathbf{Z} \, d\boldsymbol{\theta} \\
&\geqslant \iint q_{\mathbf{w}}(\mathbf{Z}, \boldsymbol{\theta}) \ln \frac{p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\theta})}{q_{\mathbf{w}}(\mathbf{Z}, \boldsymbol{\theta})} \, d\mathbf{Z} \, d\boldsymbol{\theta} \\
&= \ln p(\mathbf{X}) - \mathrm{KL}[q_{\mathbf{w}}(\mathbf{Z}, \boldsymbol{\theta}) \| p(\mathbf{Z}, \boldsymbol{\theta} | \mathbf{X})] \triangleq -\mathcal{F}(\mathbf{w}).
\end{aligned}$$

- A tractable solution is found by assuming $q_{\mathbf{w}}$ factorises given the data:

$$q_{\mathbf{w}}(\mathbf{Z}, \boldsymbol{\theta}) = \prod_n q(\mathbf{z}_n; \mathbf{w}_n) \times \prod_m q(\boldsymbol{\theta}_m; \mathbf{w}_m).$$

(Beal, 2000)

amazon

Stochastic Variational Inference

$$\mathbf{w}_n \leftarrow \arg\max_{\mathbf{w}_n} \quad \langle \ln p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\theta}) \rangle - \mathrm{KL}\left[q(\mathbf{z}_n; \mathbf{w}_n) \| p(\mathbf{z}_n)\right],$$

$$\mathbf{w}_m \leftarrow \arg\max_{\mathbf{w}_m} \quad \sum_n \langle \ln p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\theta}) \rangle - \mathrm{KL}\left[q(\boldsymbol{\theta}_m; \mathbf{w}_m) \| p(\boldsymbol{\theta}_m)\right].$$

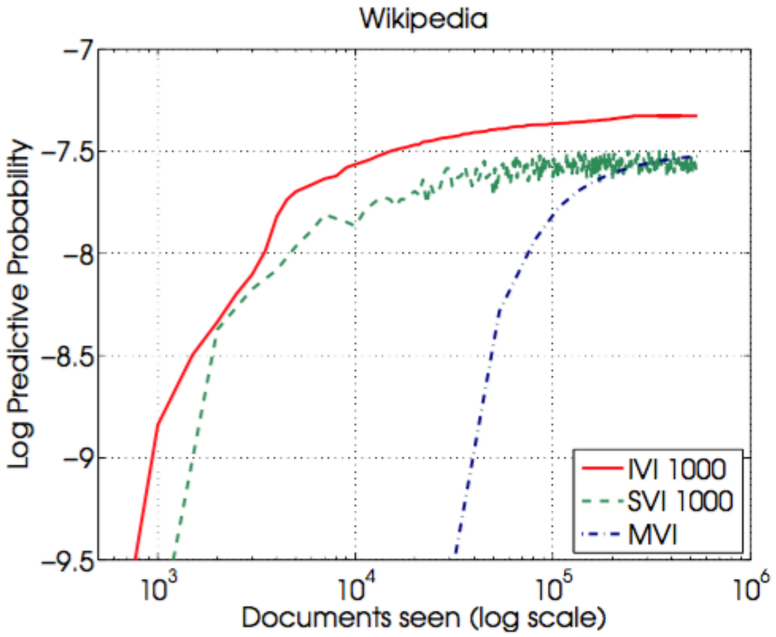
Let $\ell_n(\mathbf{w}) = \langle \ln p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\theta}) \rangle$:

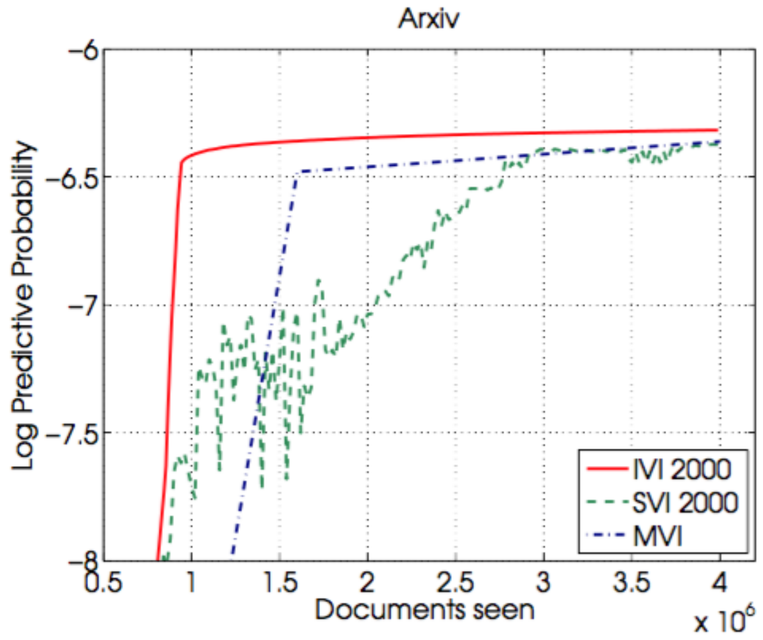
$$\mathbf{w}_m \leftarrow \mathbf{w}_m + \rho_t N \frac{\partial}{\partial \mathbf{w}_m} \left(\ell_n(\mathbf{w}) - \frac{\mathrm{KL}\left[q(\boldsymbol{\theta}_m; \mathbf{w}_m) \| p(\boldsymbol{\theta}_m)\right]}{N} \right),$$

where $\sum_t \rho_t = \infty$ and $\sum_t \rho_t^2 < \infty$.

(Hoffman, et al., NIPS 2010)

Incremental Variational Inference



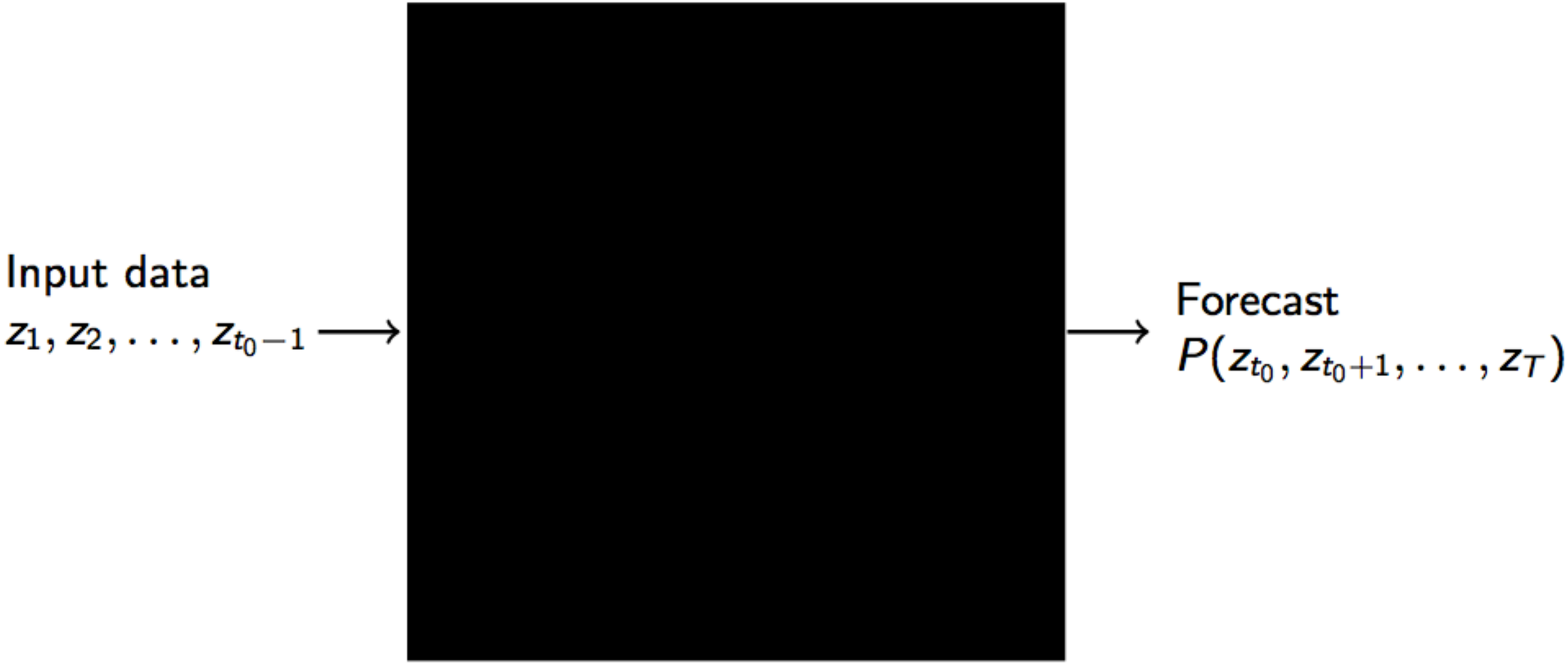


(Hinton & Neal, 1998; Schmidt, et al., 2013; Archambeau & Ermis, AABI@NIPS 2015)

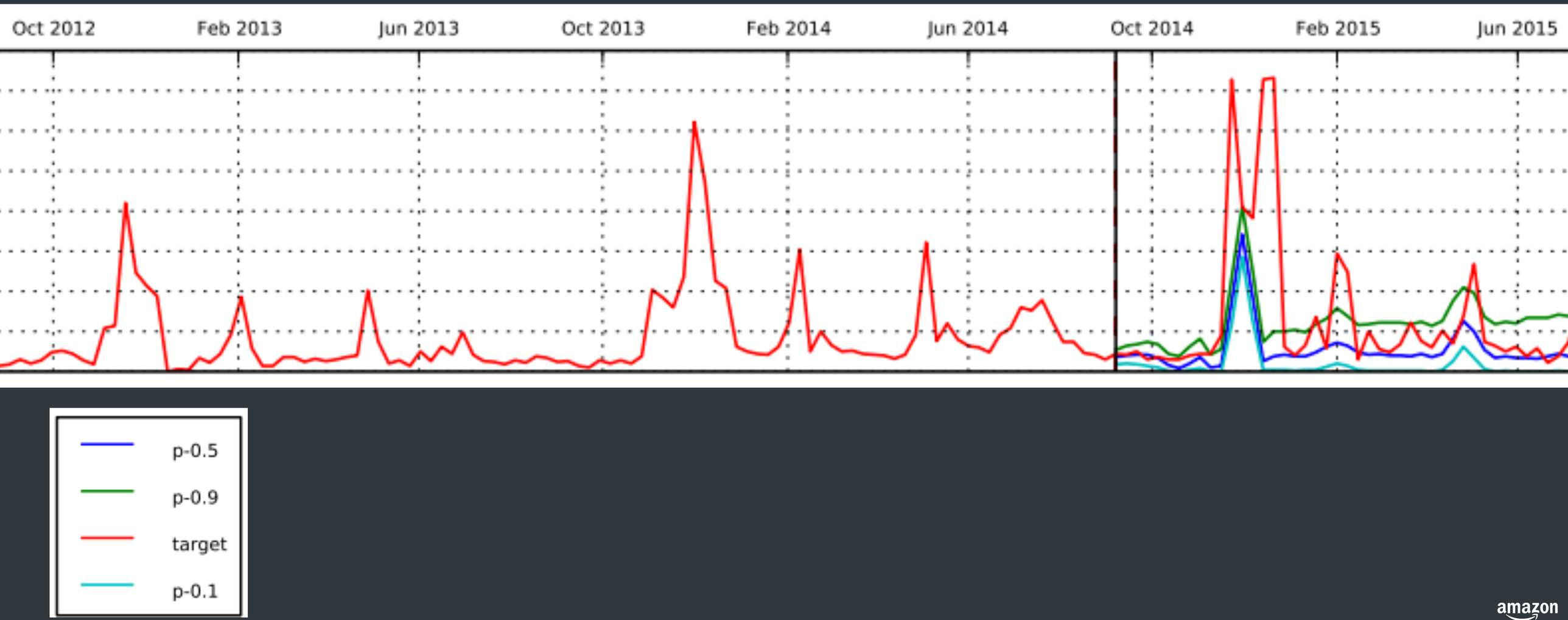
Demand Forecasting

amazon

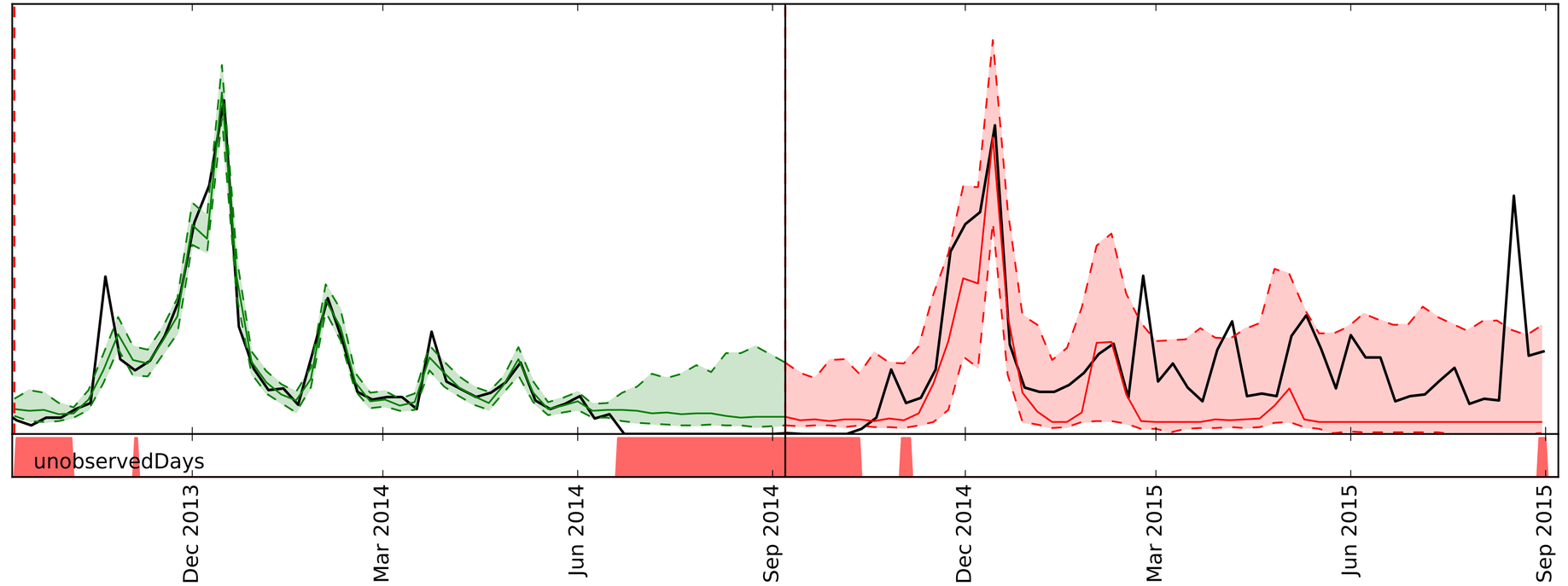
Forecasting

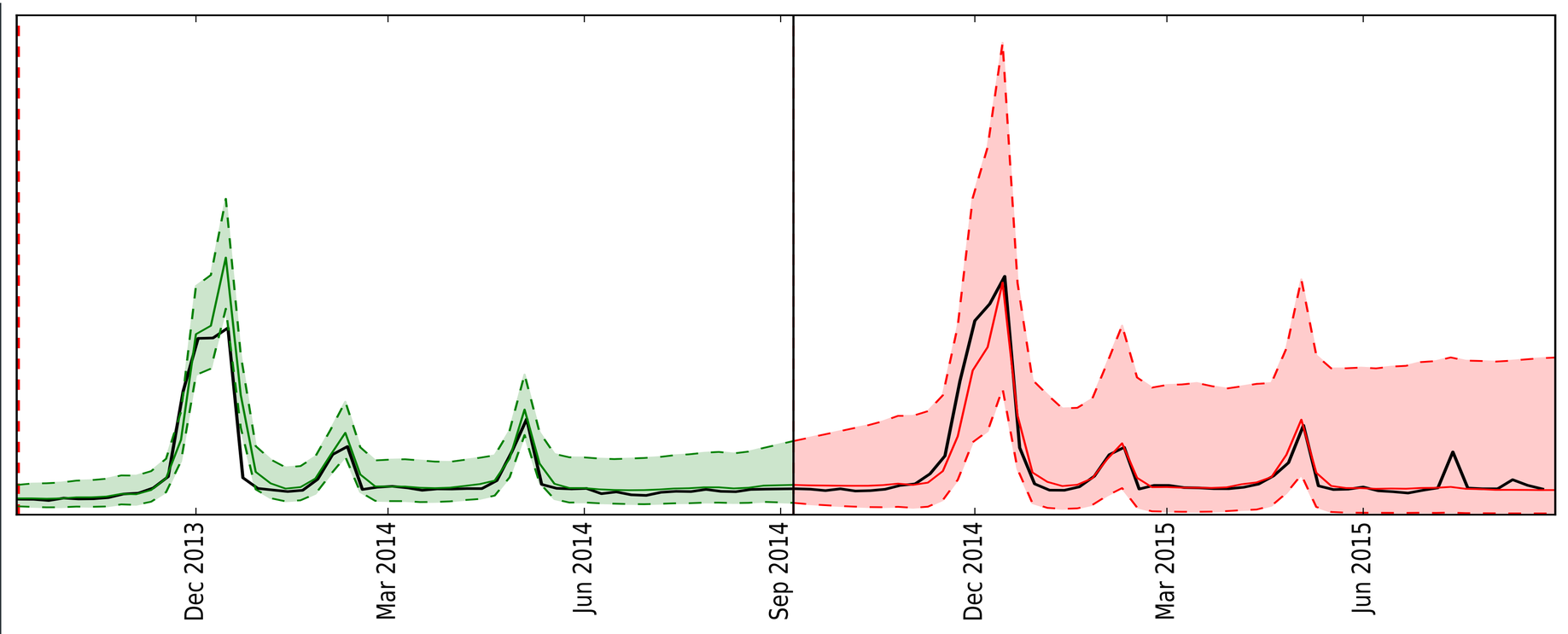


Seasonality and External Events



amazon





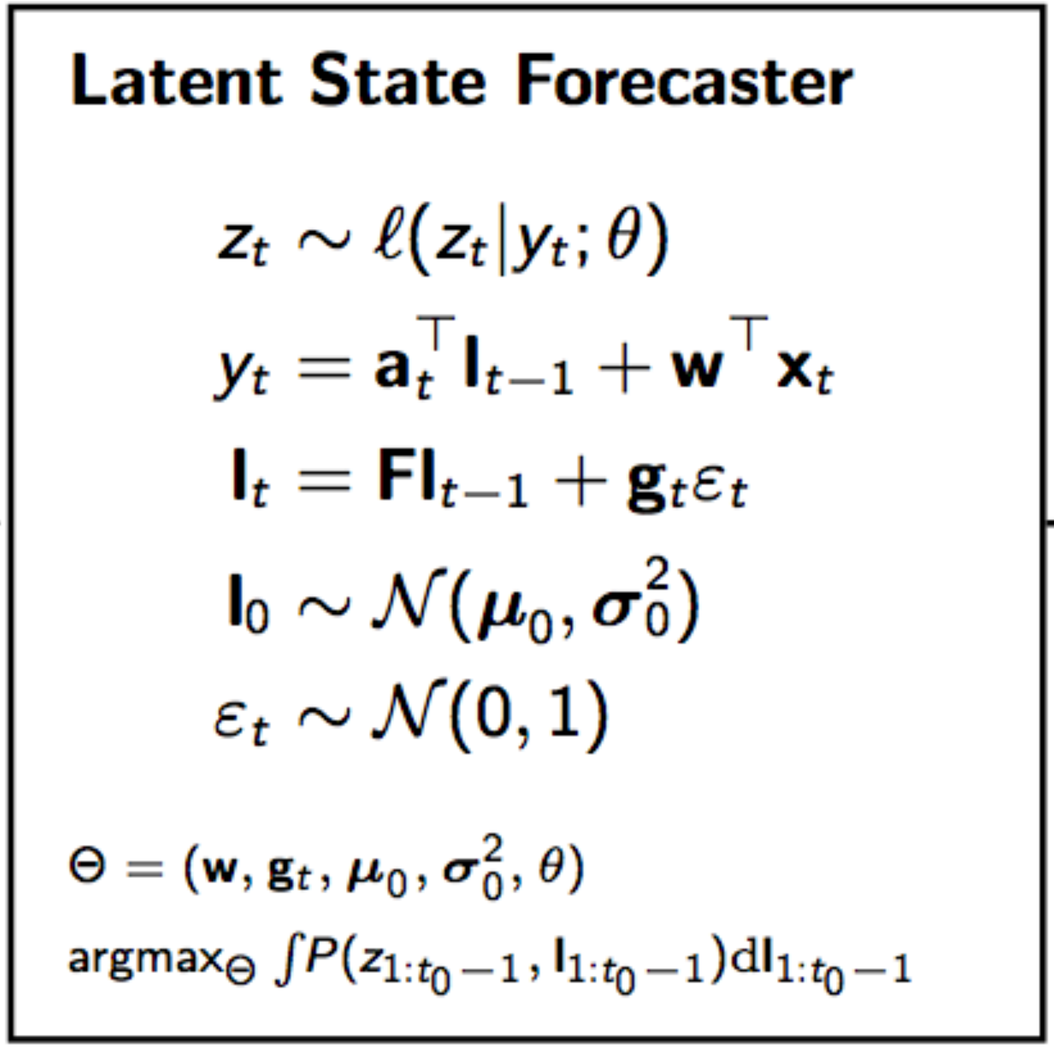
Probabilistic Forecasts

amazon

Input data

$z_1, z_2, \ldots, z_{t_0-1}$

$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$



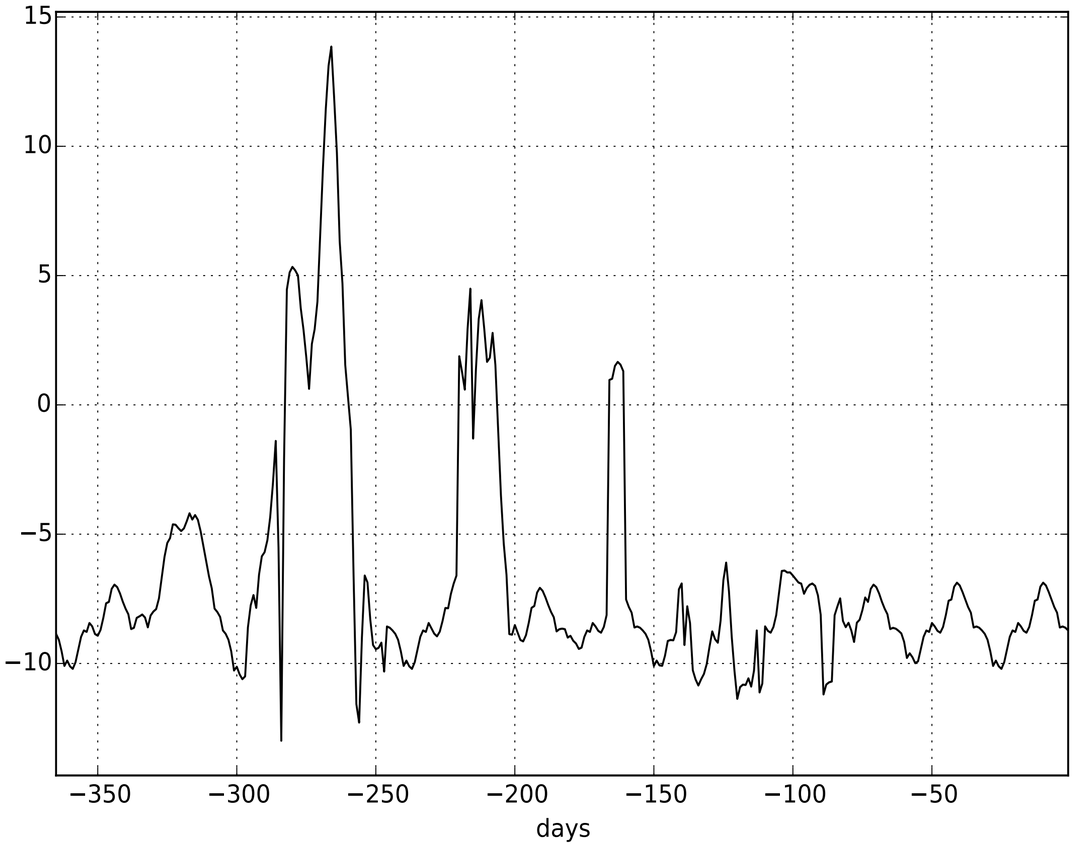
Forecast

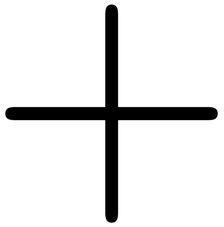
$P(z_{t_0}, z_{t_0+1}, \ldots, z_T)$

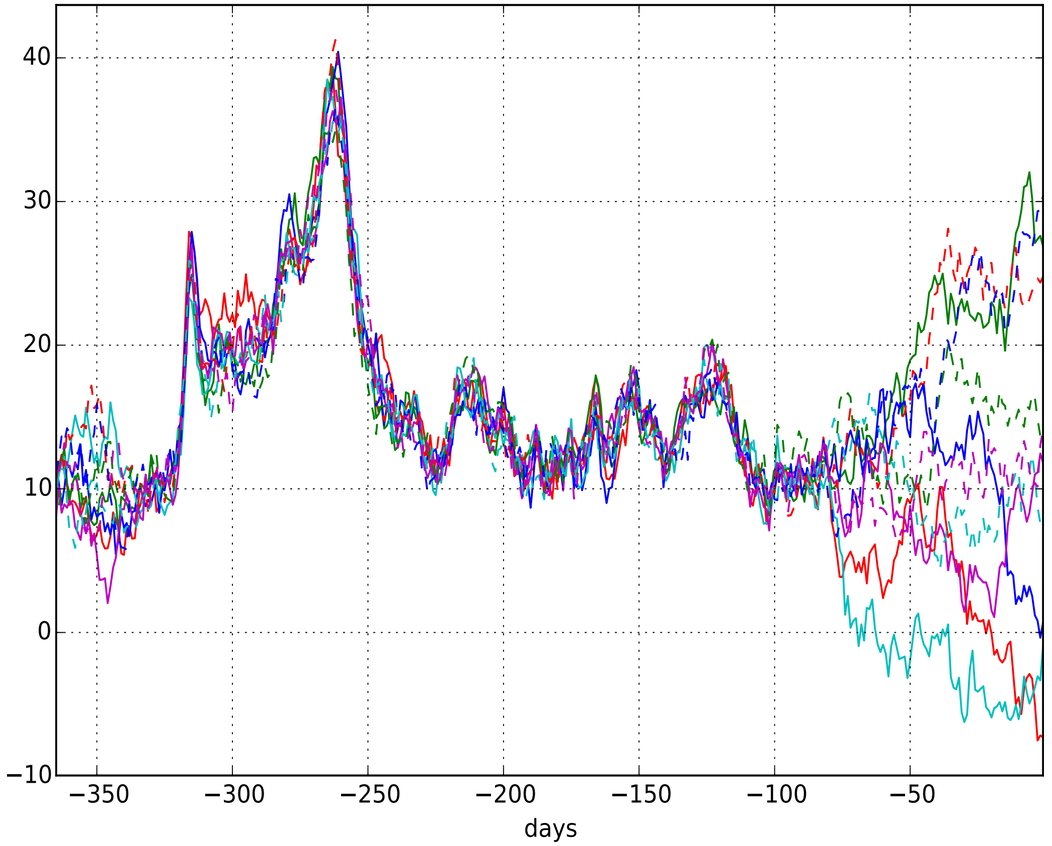
(Seeger, et al.: Bayesian Intermittent Demand Forecasting for Large Inventories. NIPS 2016.)

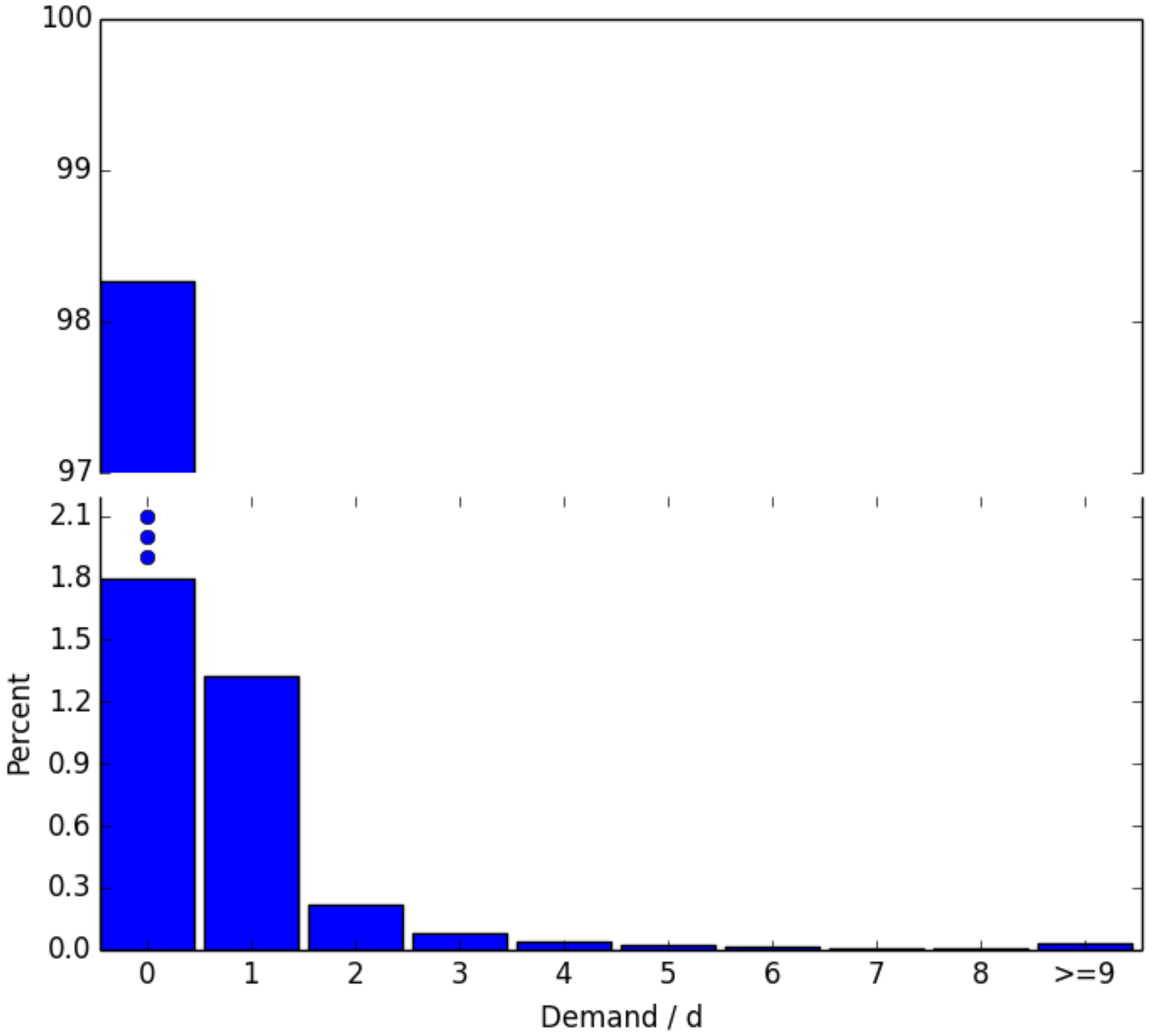
amazon

Latent Gaussian State



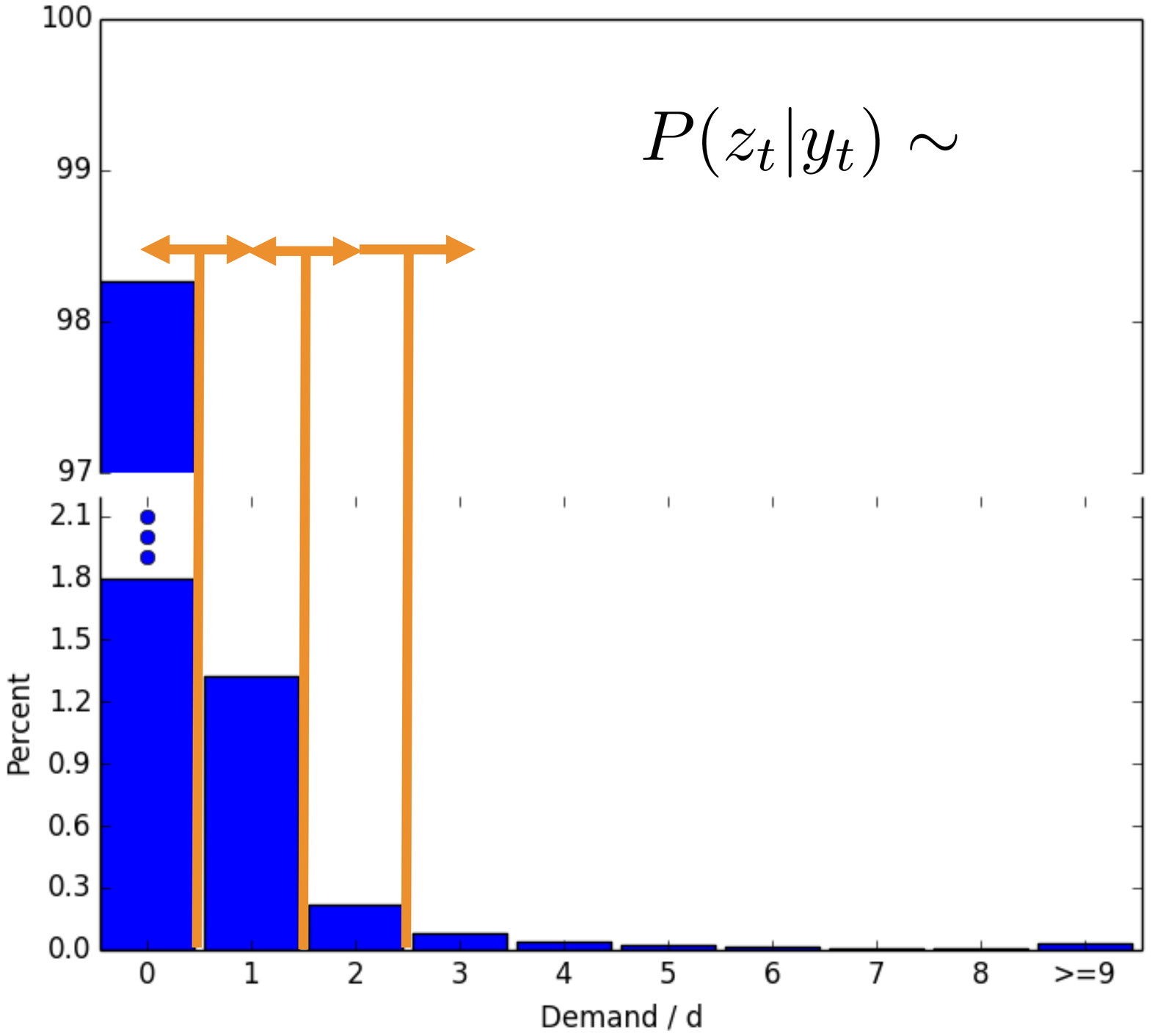






Intermittent Demand

amazon



Multi-stage Likelihood

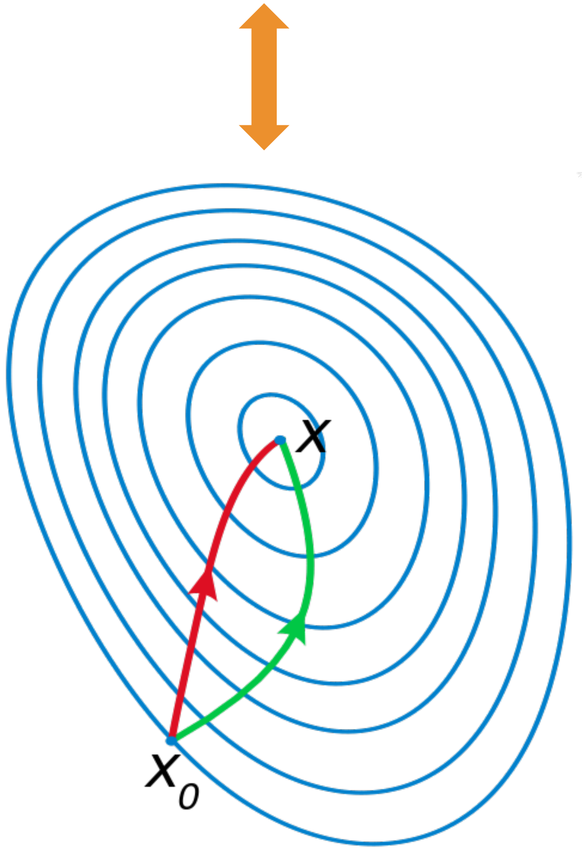
amazon

Scalable Inference

```
while (!converged) {

    b = kalmanSmoothing(a)

    a = glueCode(b)

}
```

$$\phi(\theta) = -\log \int P(\mathbf{z}_{\text{train}}, \mathbf{s} \,|\, \theta)\, d\mathbf{s}$$

$$[\Phi(\theta), \nabla_\theta \Phi]$$



amazon

	Parts	EC-sub	EC-all
# items	19874	39700	534884
Unit t	month	day	day
Median CV^2	2.4	5.8	9.7
Freq. $z_t = 0$	54%	46%	83%
In-stock ratio	100%	73%	71%
Avg. size series	33	329	293
# item-days	656K	13M	157M

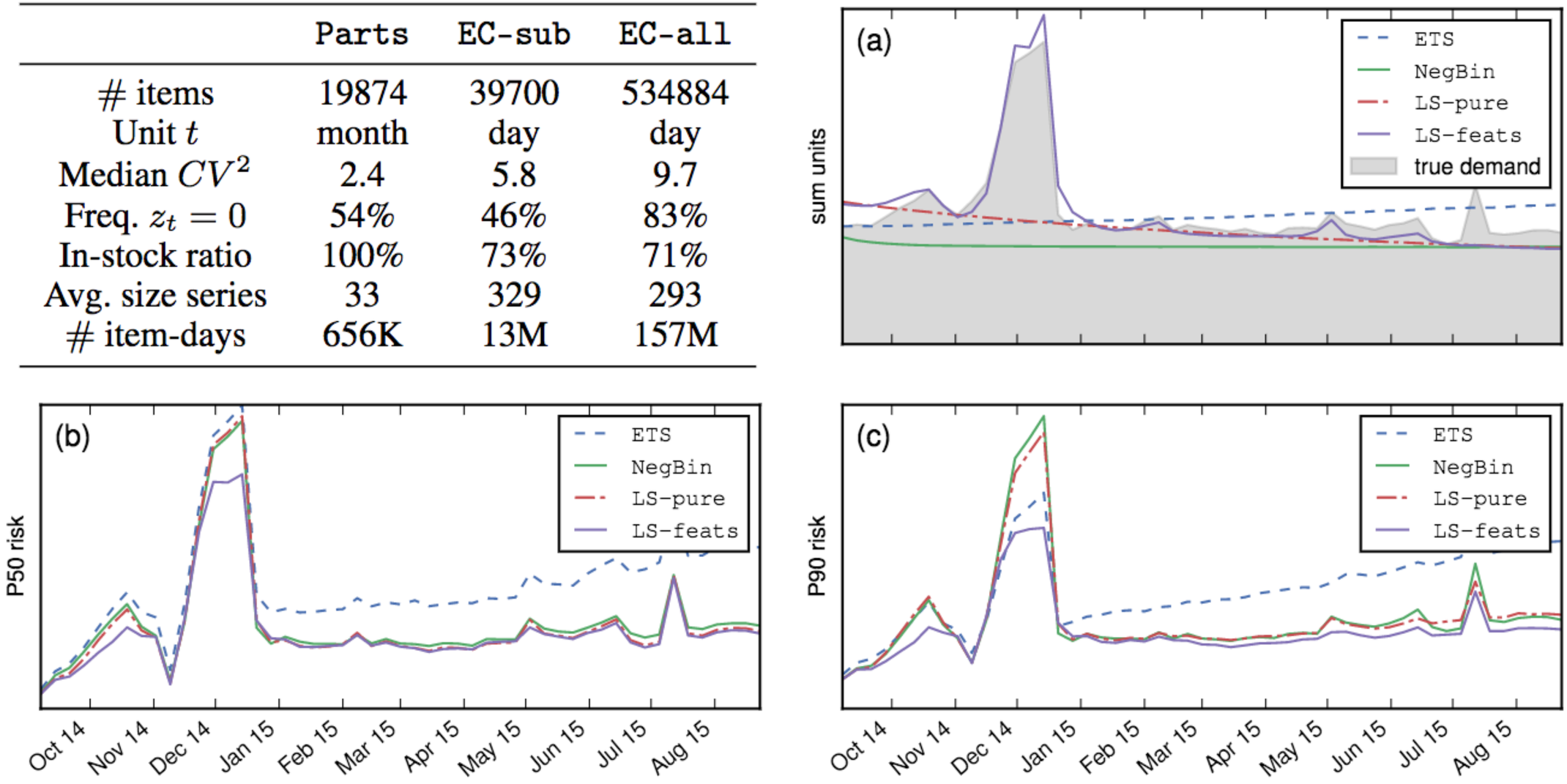


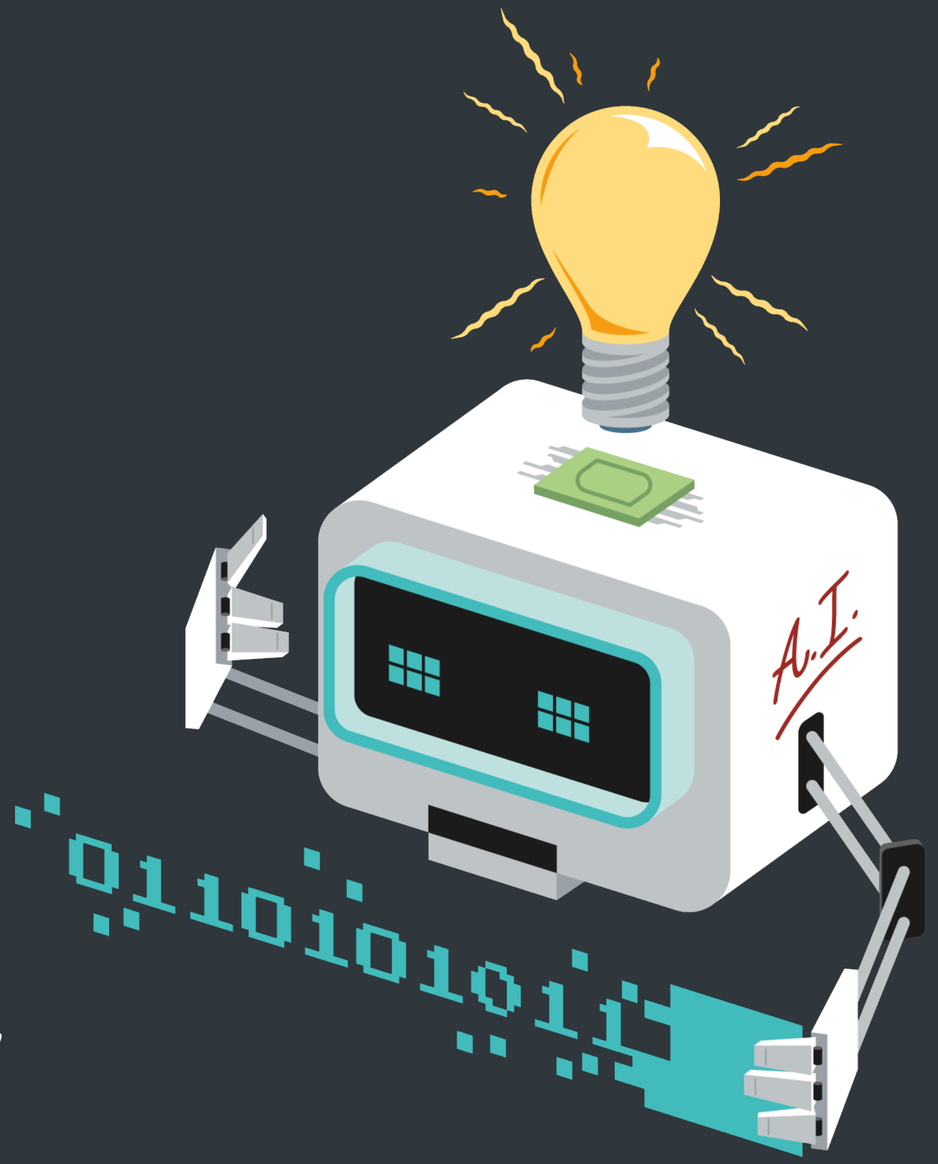
Figure 3: **Table**: Dataset properties. $CV^2 = \mathrm{Var}[z_t]/\mathrm{E}[z_t]^2$ measures burstiness. **(a)**: Sum of weekly P50 point (median) forecast over a one-year prediction range for the different methods (lines) as well as sum of true demand (shaded area), on dataset $\mathcal{I} =$ EC-sub. **(b)**: Weekly P50 risk $R^{0.5}[\mathcal{I}; (7 \cdot k, 7)]$, $k = 0, 1, \ldots,$ for same dataset. **(c)**: Same as (b) for P90 risk.

Take away

Machine learning & AI are transformational.

Wealth of probabilistic machine learning applications.

When deploying machine learning, keep it simple (not simpler).



amazon

Beyza Ermis

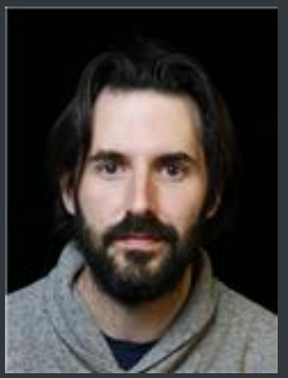
Felix Bießmann

Michael Brückner

Matthias Seeger

David Salinas

Valentin Flunkert



Pablo García Moreno

amazon

cedrica@amazon.com