User-friendly conjugacy for completely random measures

Tamara Broderick, Michael I. Jordan
UC Berkeley

Conjugacy

Conjugacy

(Nonparametric) Bayesian concerns:

- Calculating the posterior can be hard

Conjugacy

(Nonparametric) Bayesian concerns:

- Calculating the posterior can be hard
- How to choose the prior

Conjugacy

(Nonparametric) Bayesian concerns:

- Calculating the posterior can be hard
- How to choose the prior

One solution: Conjugacy

Conjugacy

(Nonparametric) Bayesian concerns:

- Calculating the posterior can be hard
- How to choose the prior

One solution: Conjugacy

- Can be (sometimes unnecessarily) restrictive in parametric setting

Conjugacy

(Nonparametric) Bayesian concerns:

- Calculating the posterior can be hard
- How to choose the prior

One solution: Conjugacy

- Can be (sometimes unnecessarily) restrictive in parametric setting
- Computational imperative in nonparametric setting

Conjugacy

(Nonparametric) Bayesian concerns:

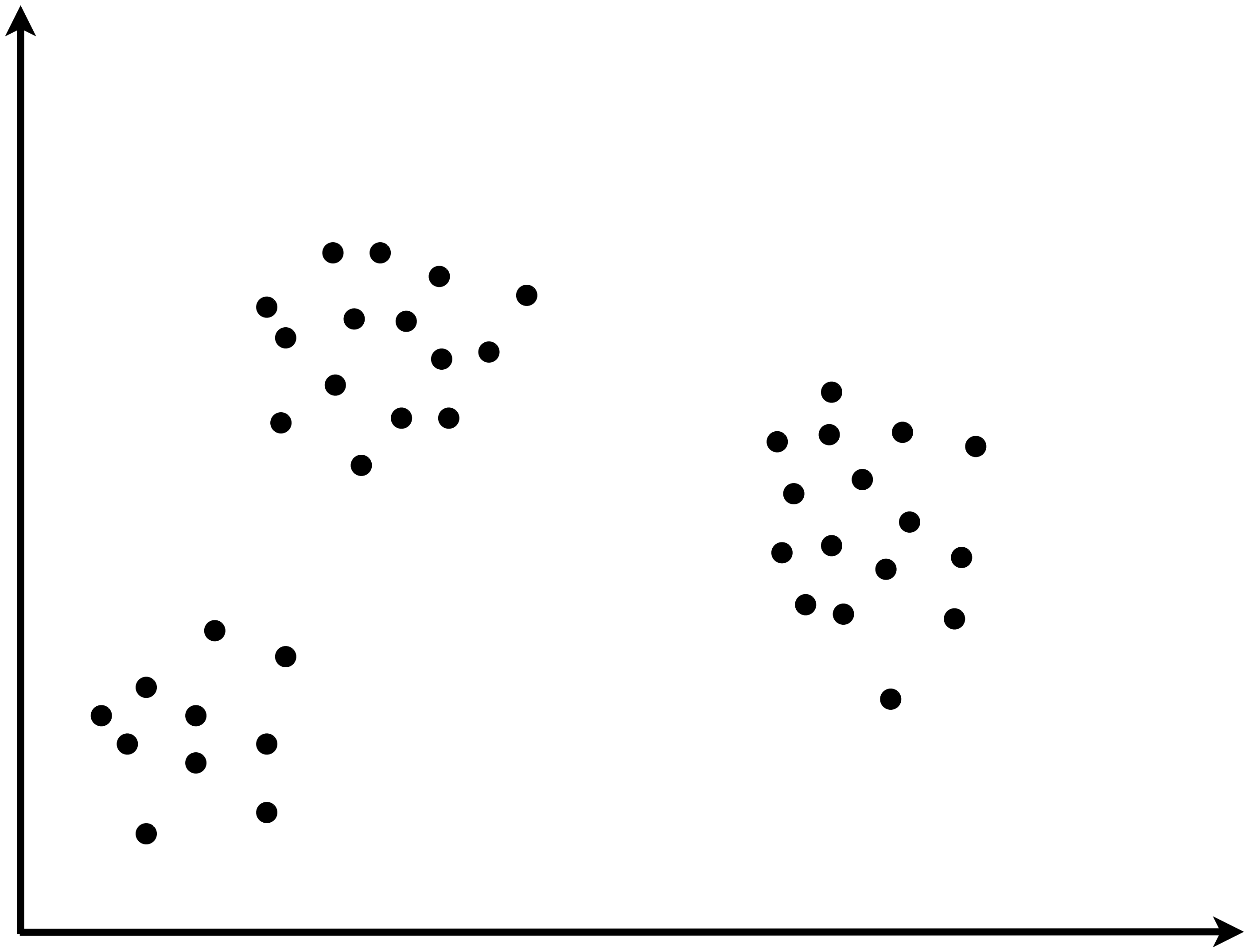
- Calculating the posterior can be hard
- How to choose the prior

One solution: Conjugacy

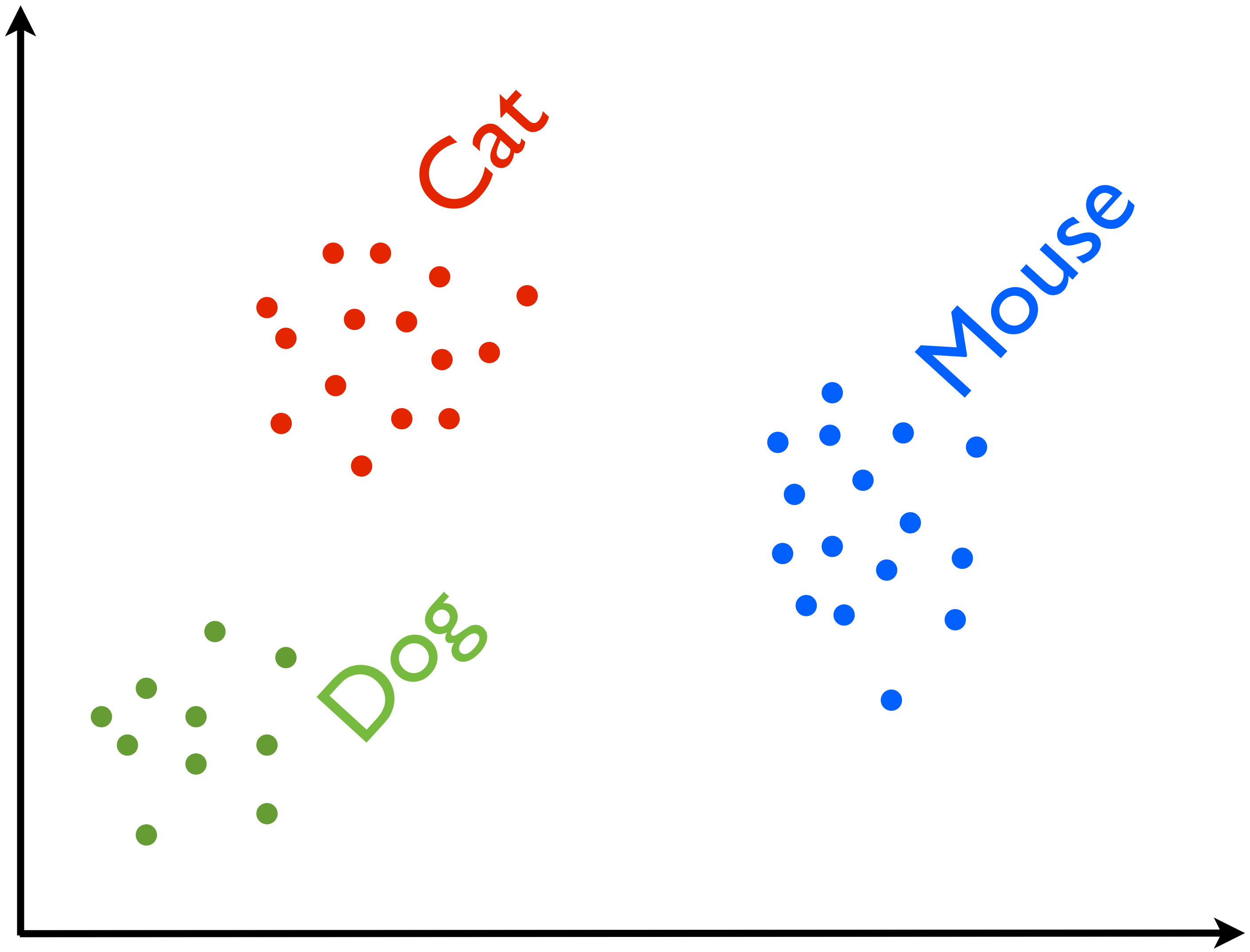
- Can be (sometimes unnecessarily) restrictive in parametric setting
- Computational imperative in nonparametric setting

We want conjugacy to be user-friendly in Bayesian nonparametrics

Clustering

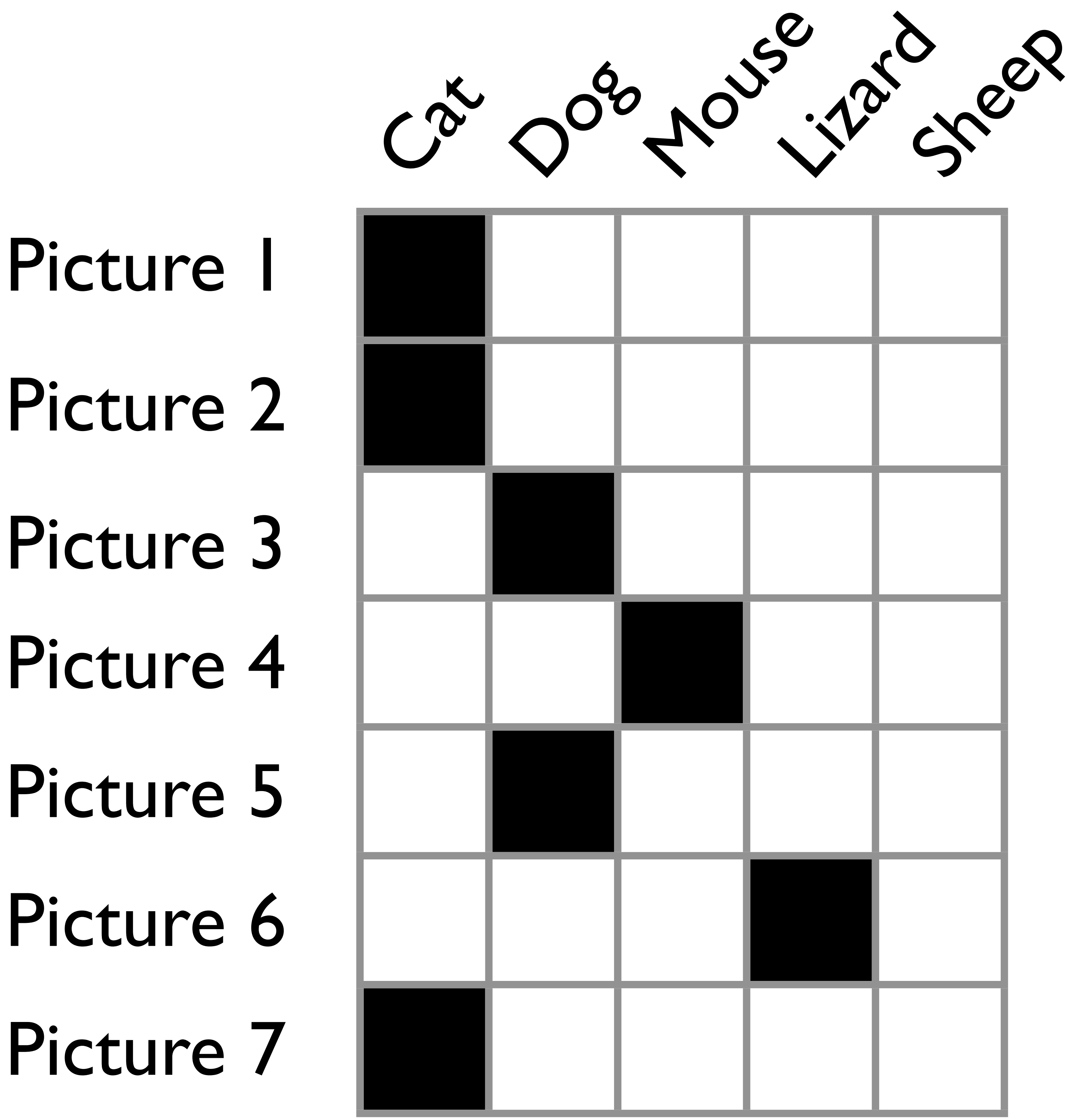


Clustering

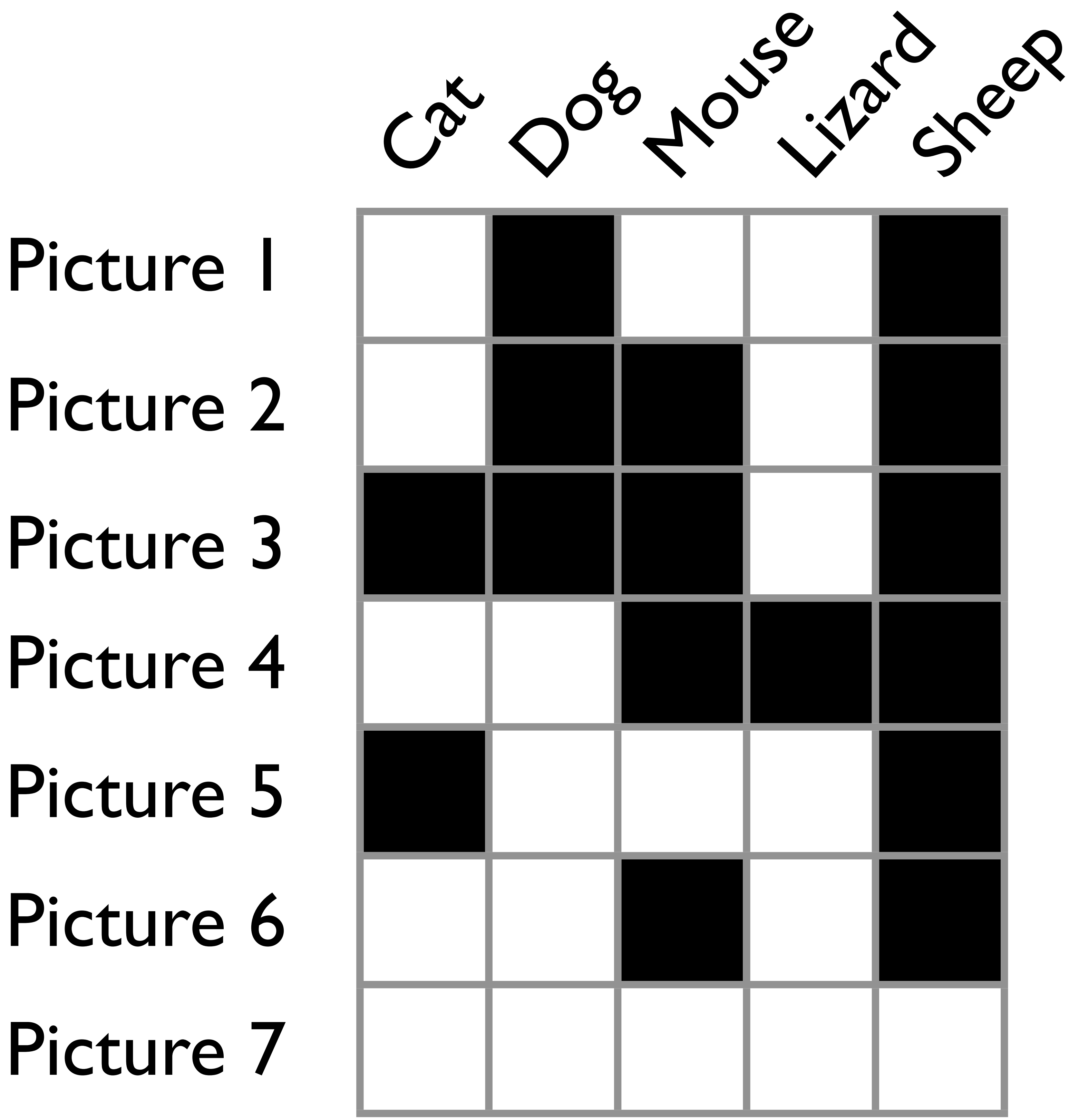


“clusters”

Clustering

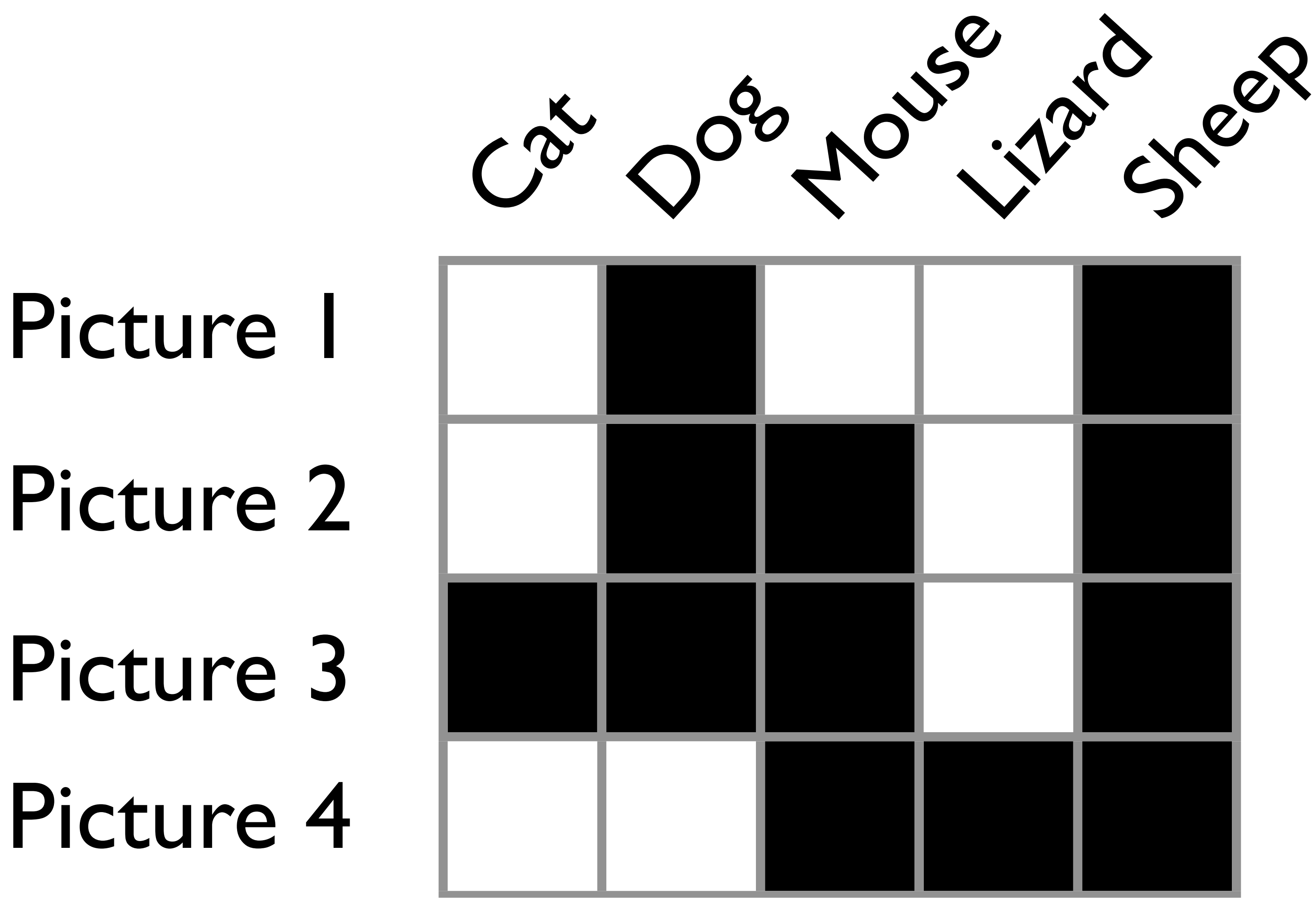

Cat
Dog
Mouse
Lizard
Sheep
Picture 1
Picture 2
Picture 3
Picture 4
Picture 5
Picture 6
Picture 7

Feature allocation



“features”

Feature allocation

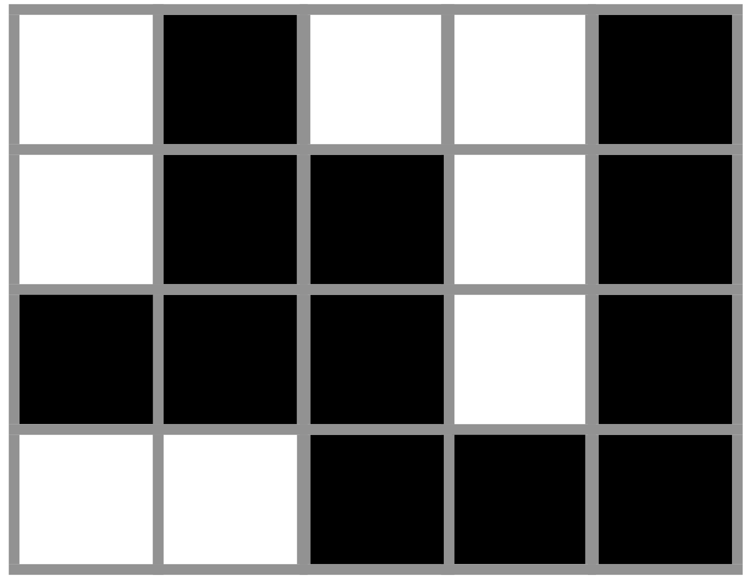


“features”

Indian buffet process [Griffiths, Ghahramani 2006]

Feature allocation

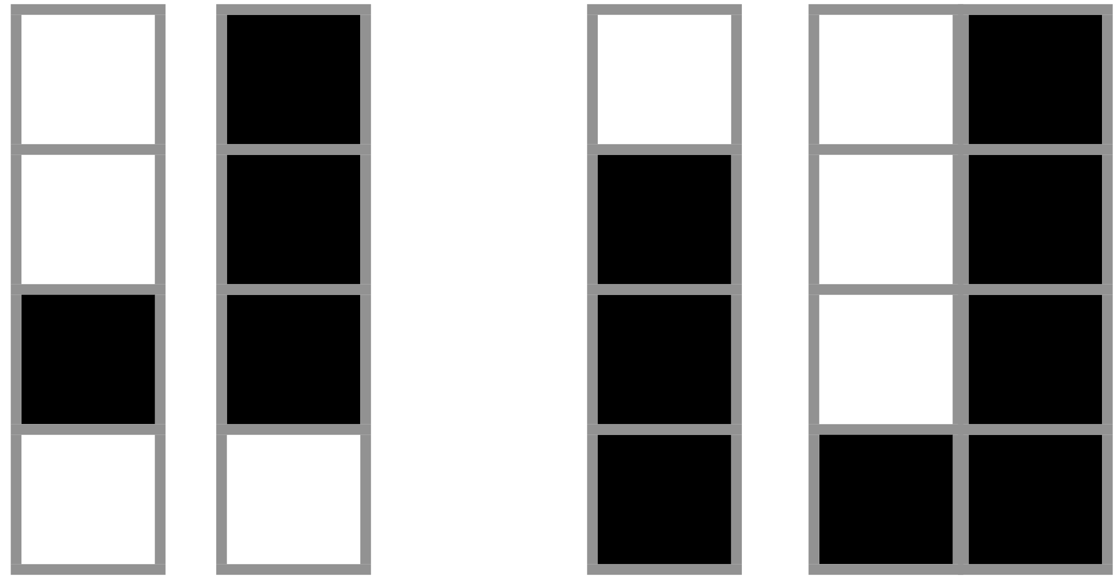
Picture 1
Picture 2
Picture 3
Picture 4



[Thibaux, Jordan 2007]

Feature allocation

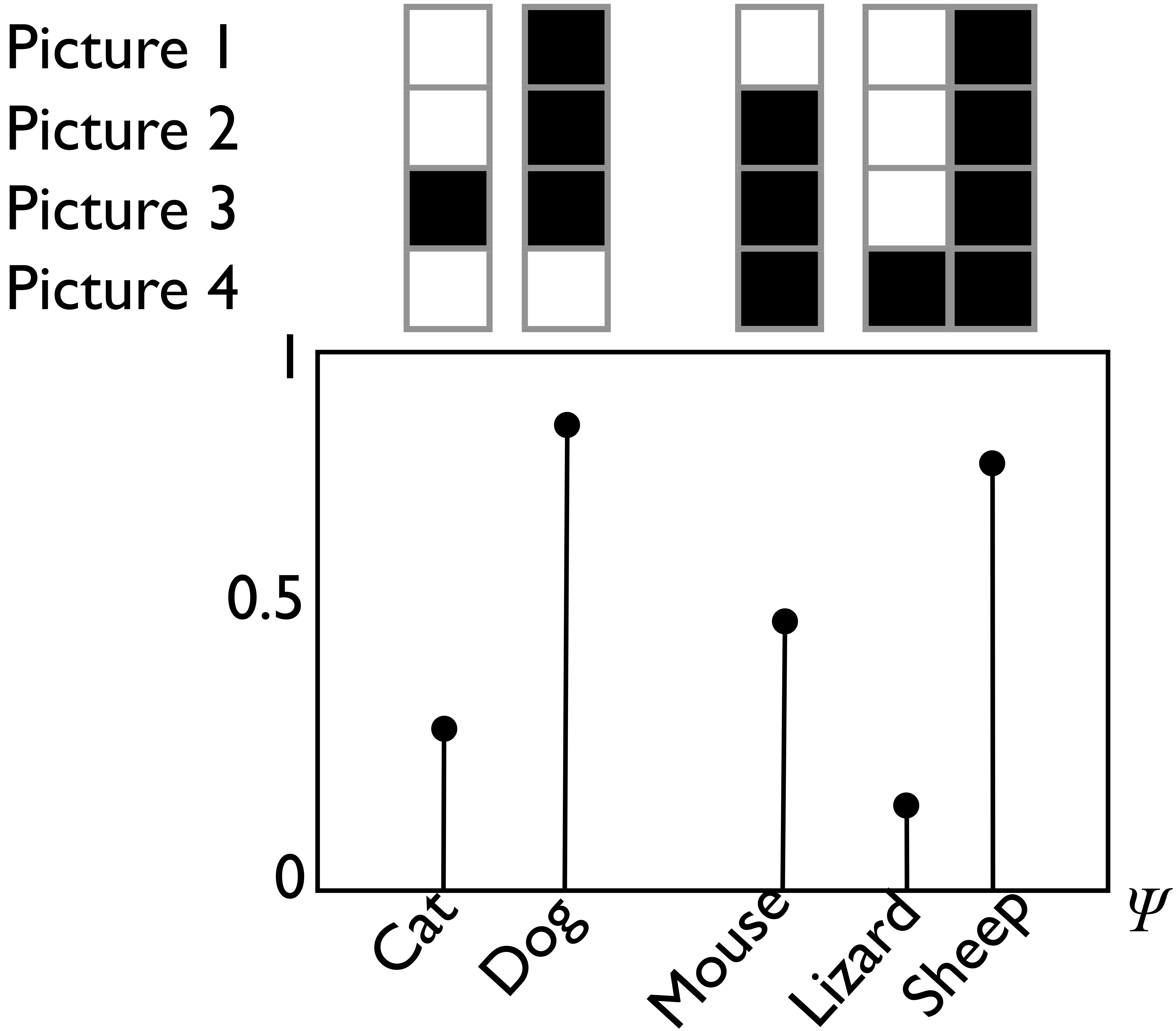
Picture 1
Picture 2
Picture 3
Picture 4



[Thibaux, Jordan 2007]

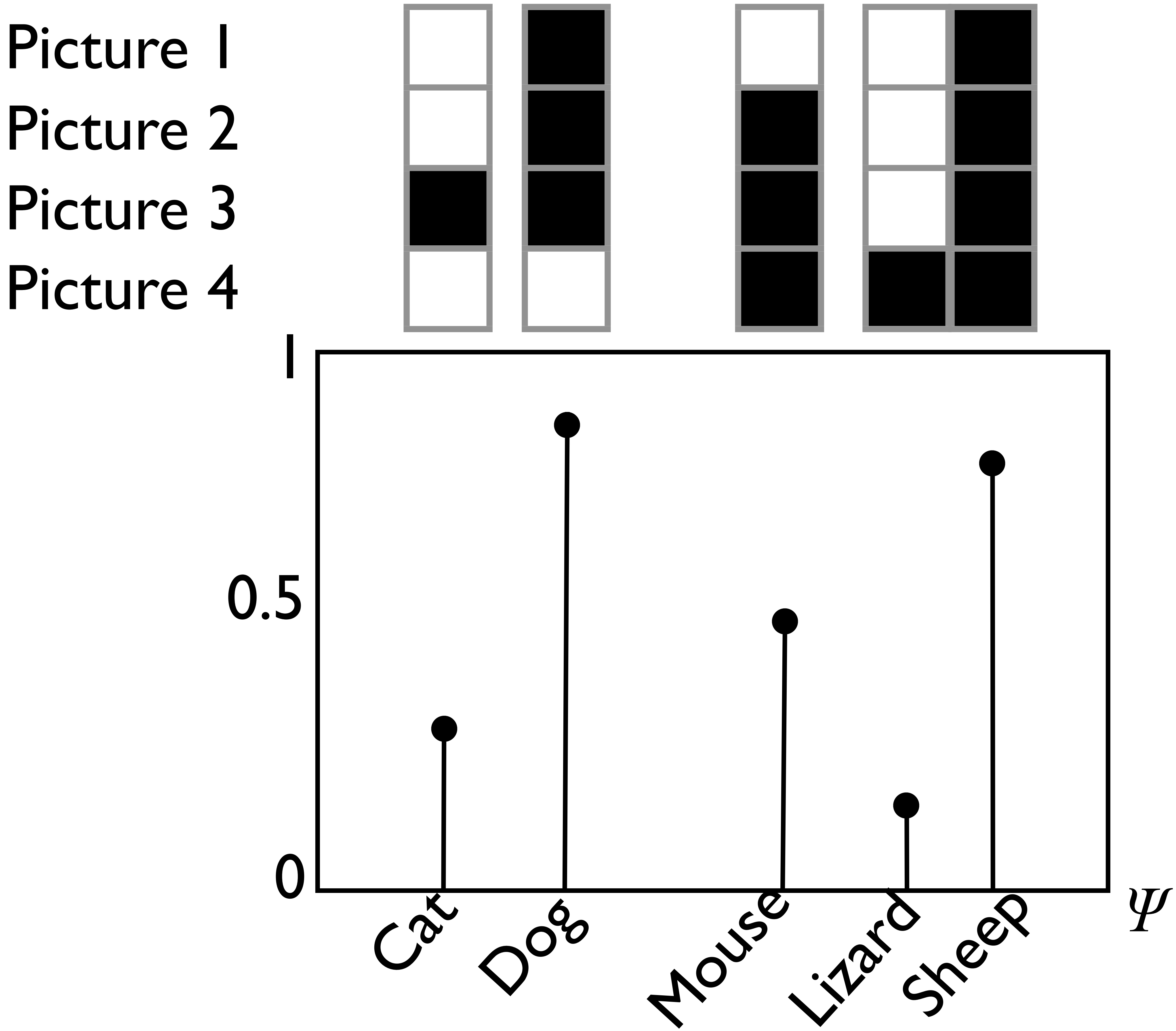
Feature allocation

[Thibaux, Jordan 2007]



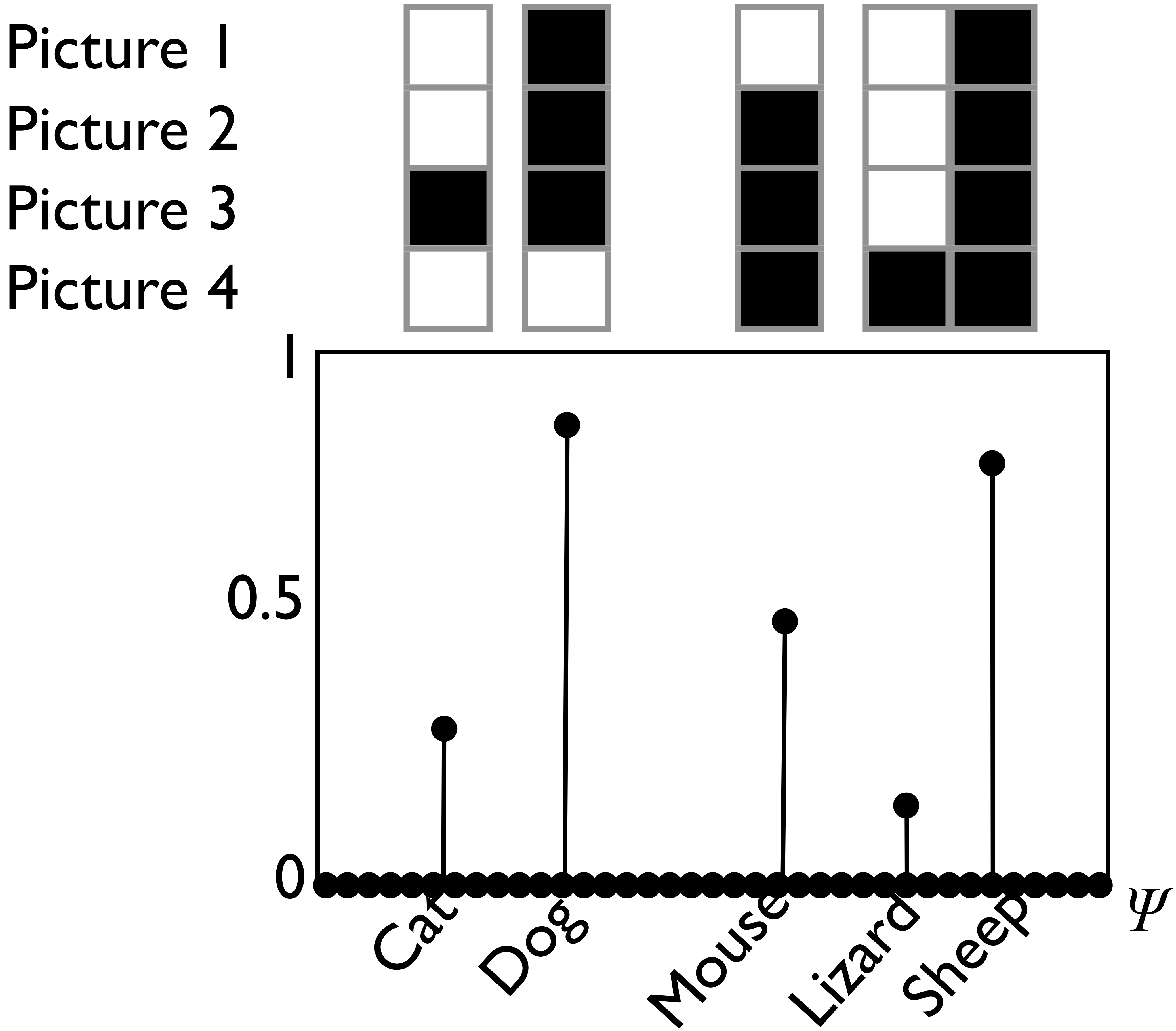
Feature allocation

[Thibaux, Jordan 2007]



"beta process"

Feature allocation

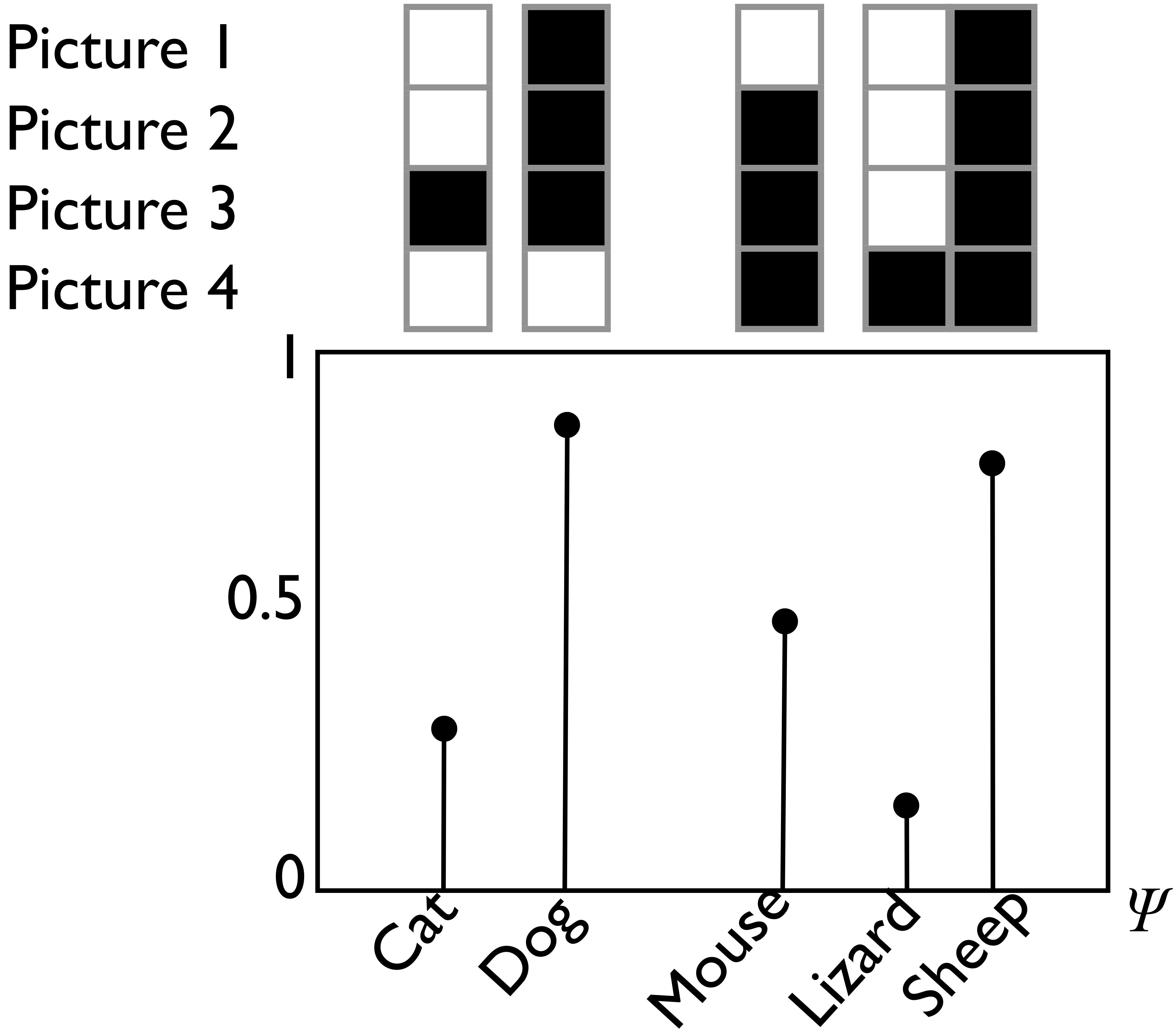


[Thibaux, Jordan 2007]

"beta process"

Feature allocation

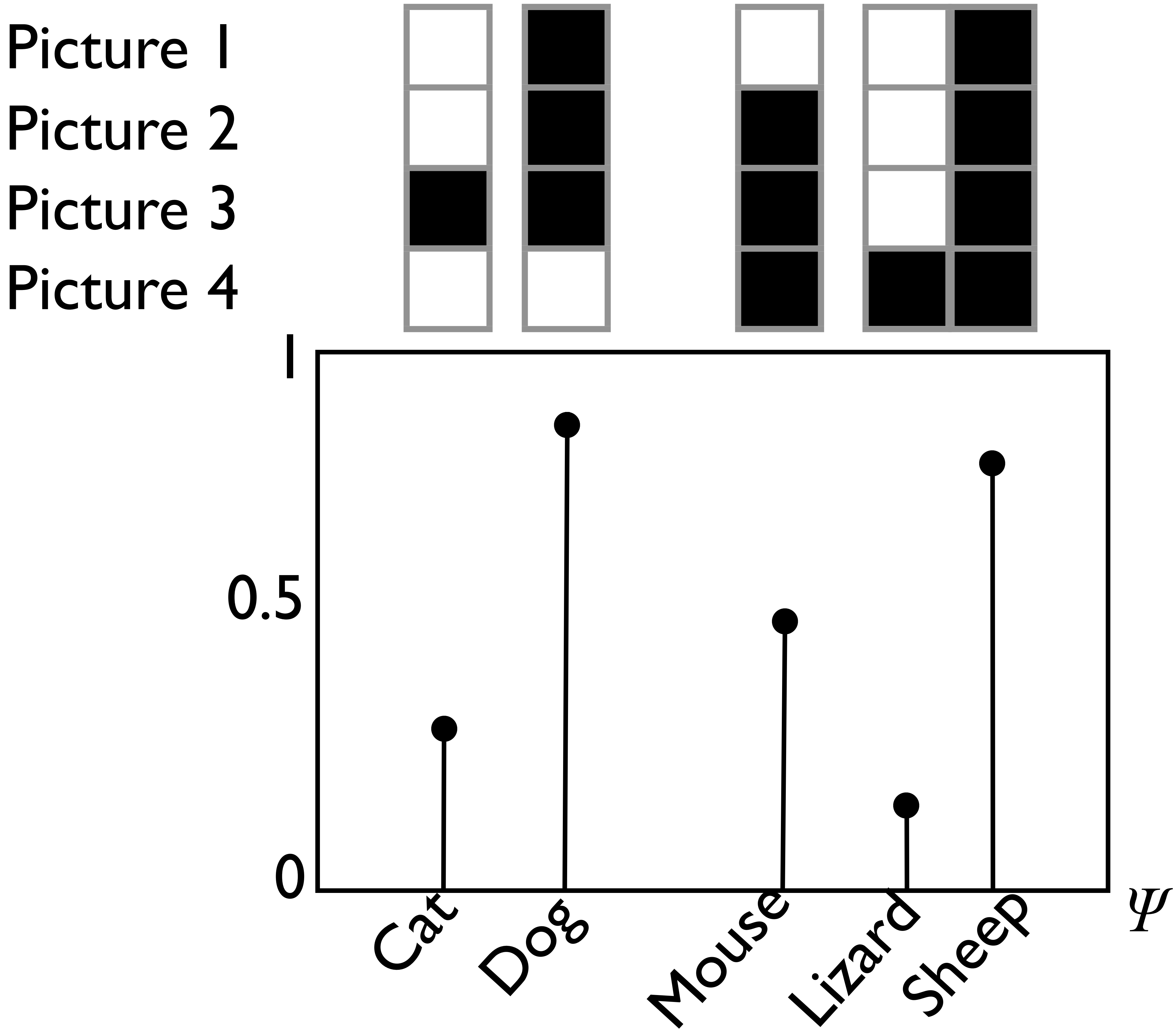
[Thibaux, Jordan 2007]



“beta process”

Feature allocation

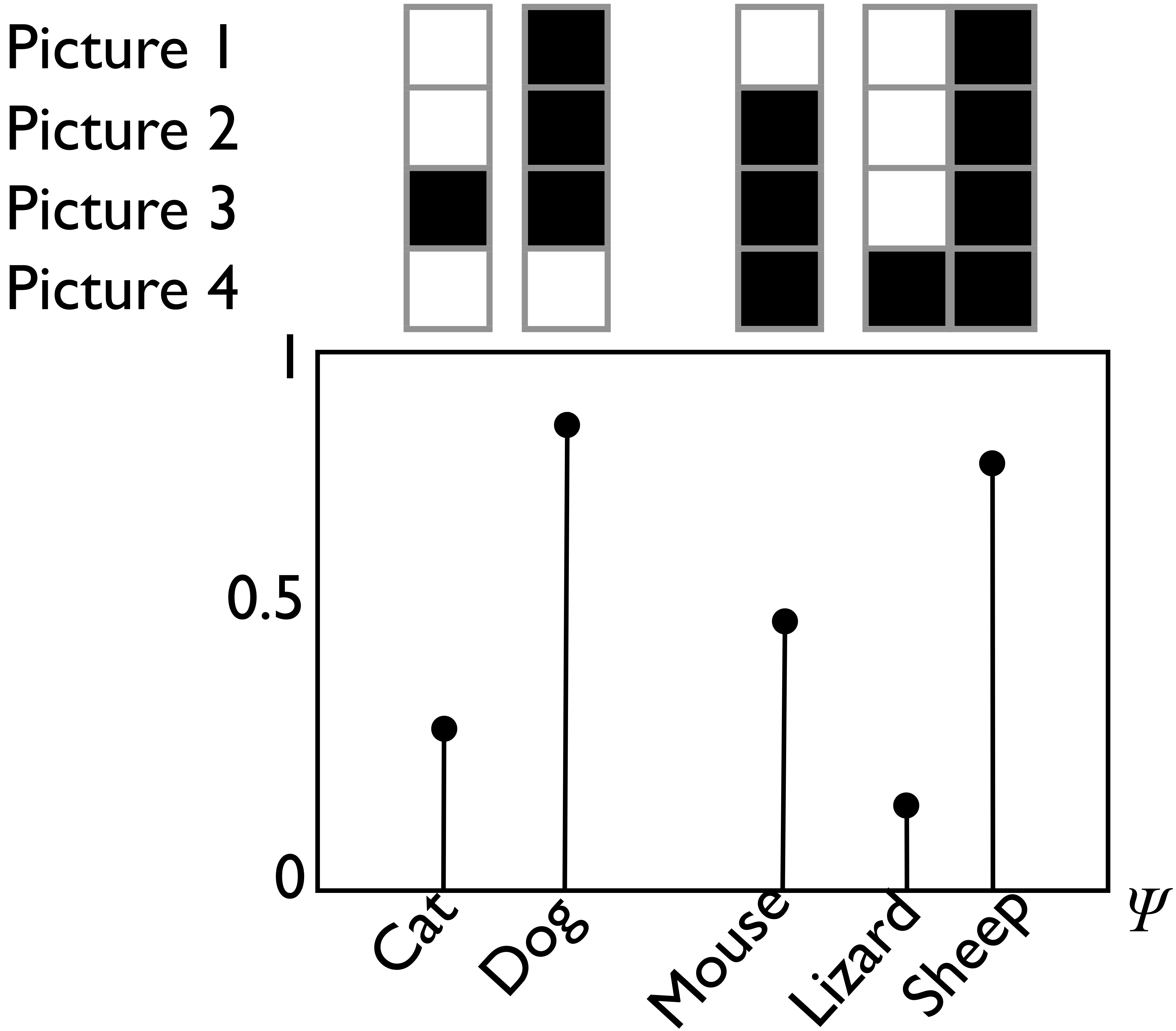
[Hjort 1990; Kim 1999; Thibaux, Jordan 2007]



"beta process"

Feature allocation

[Hjort 1990; Kim 1999; Thibaux, Jordan 2007]

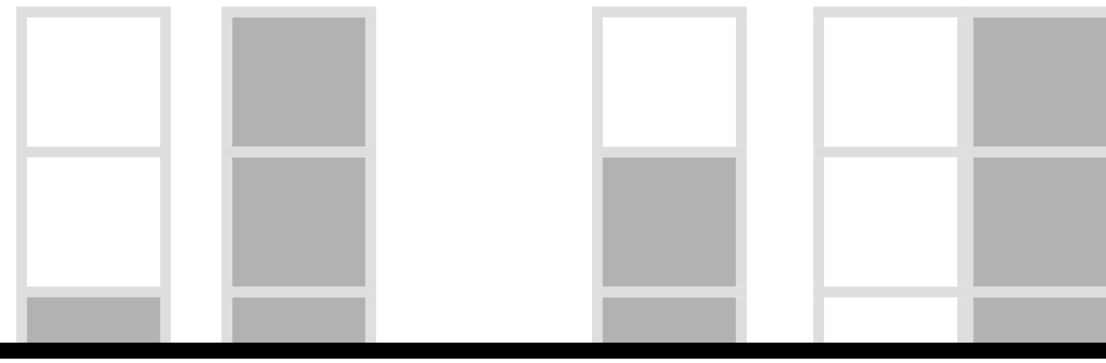


"beta process"

Feature allocation

[Hjort 1990; Kim 1999; Thibaux, Jordan 2007]

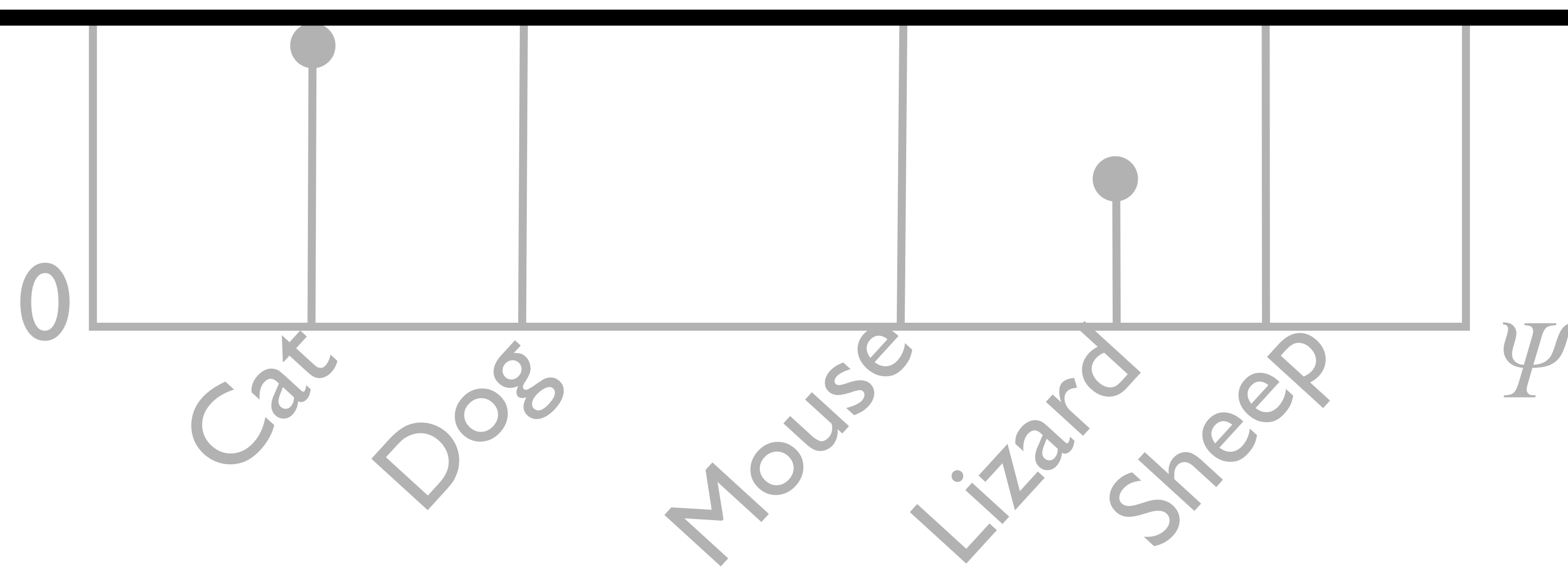
Picture 1
Picture 2
Pic
Pic



Want to show:

Posterior is a beta process

- No taking limits to infinite case
- Understandable
- Generalizable (other priors, likelihoods)



Poisson process

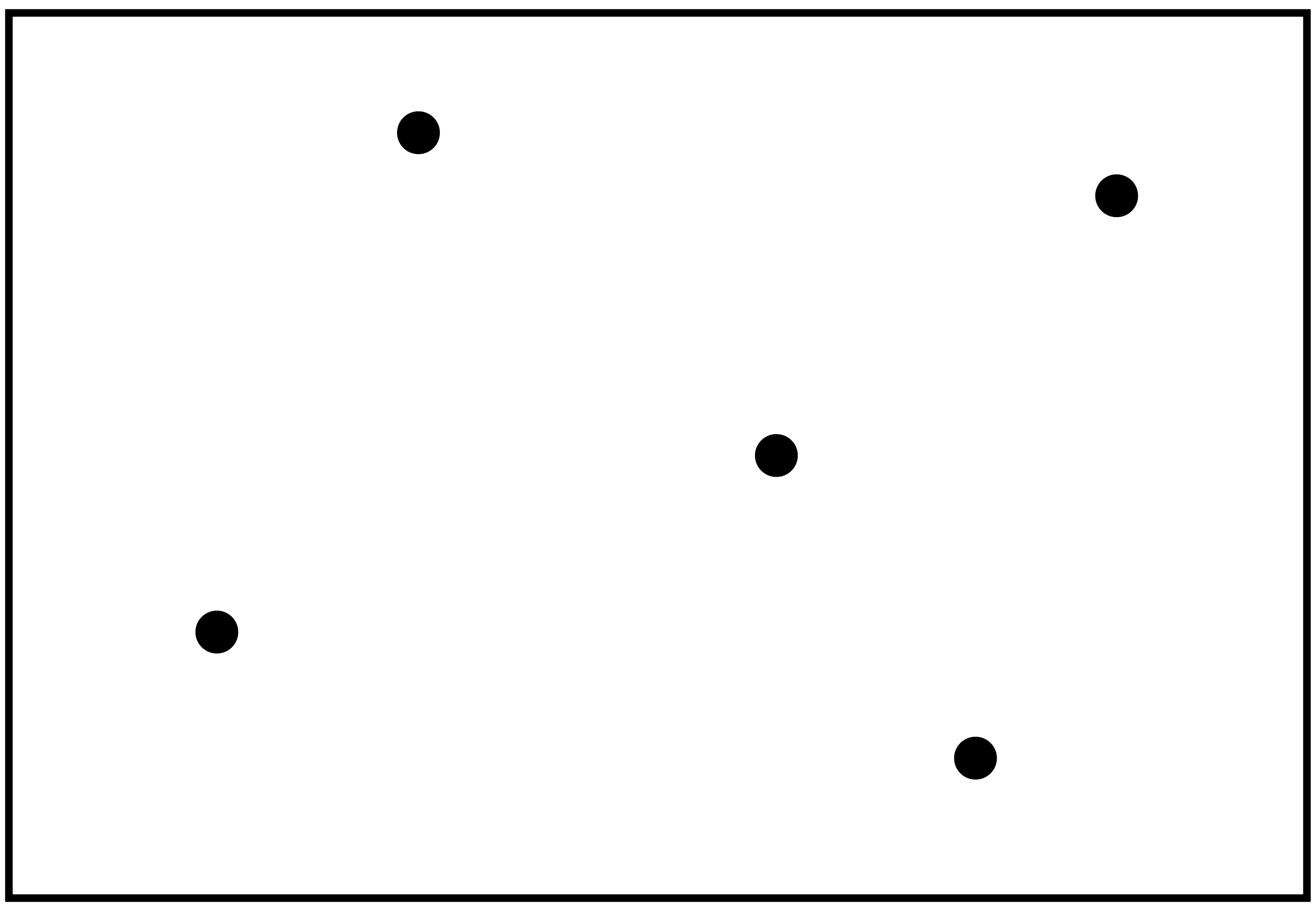
with mean measure ν

[Kingman 1993]

Poisson process

with mean measure ν

Random, countable set of points with:

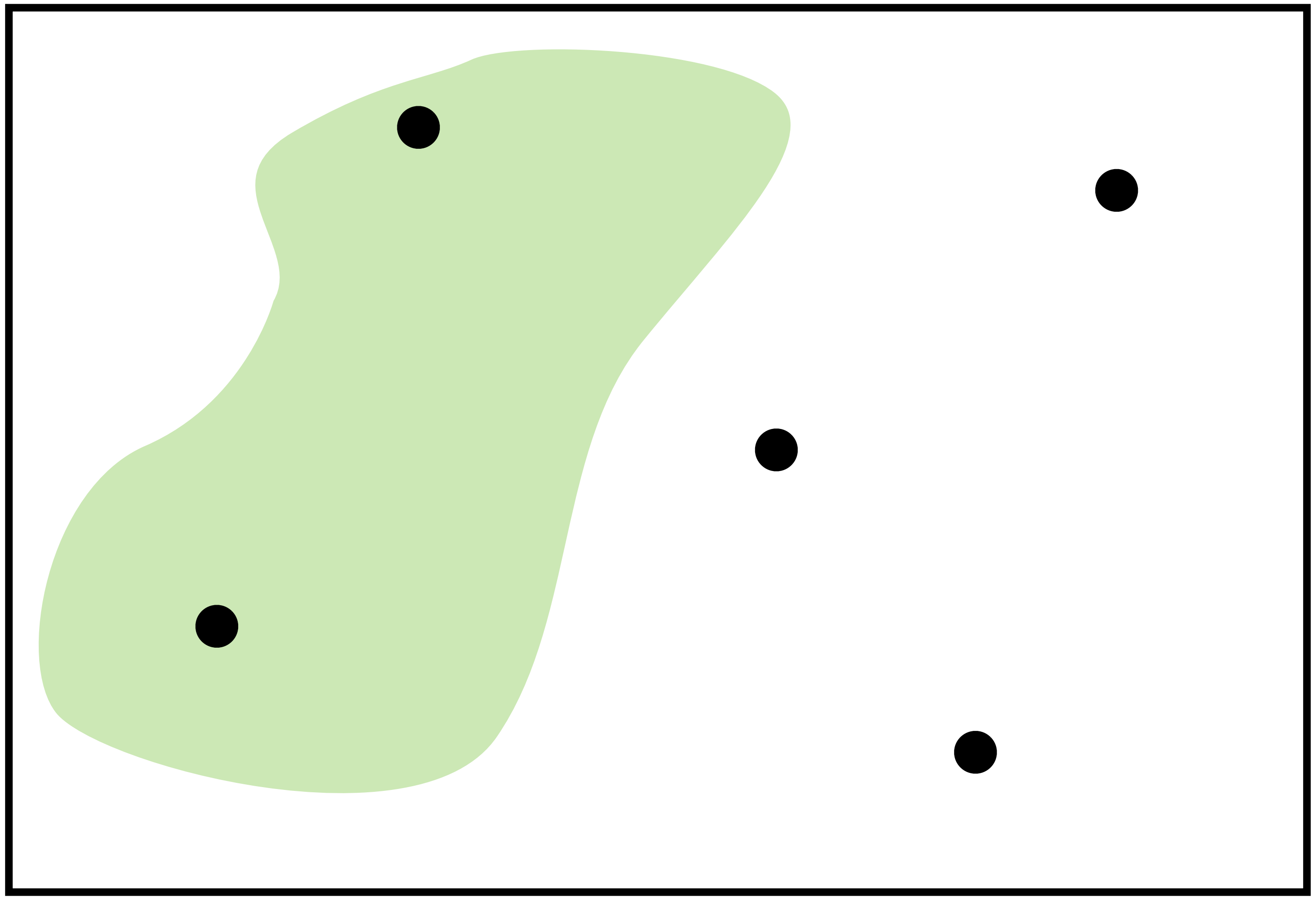


[Kingman 1993]

Poisson process
with mean measure ν

Random, countable set of points with:

- # points in set *A* is $\sim \mathrm{Poisson}[\nu(A)]$



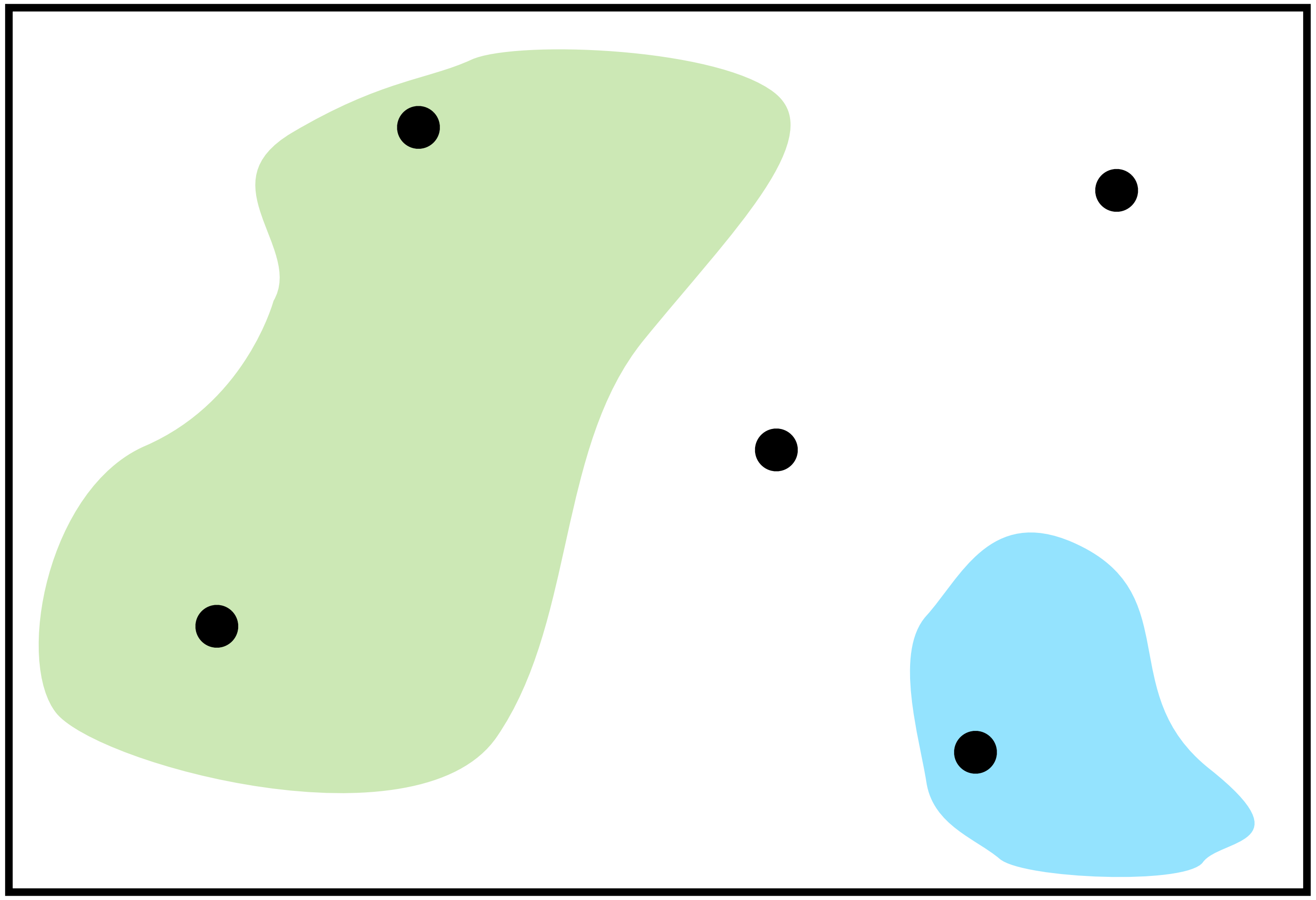
[Kingman 1993]

Poisson process

with mean measure ν

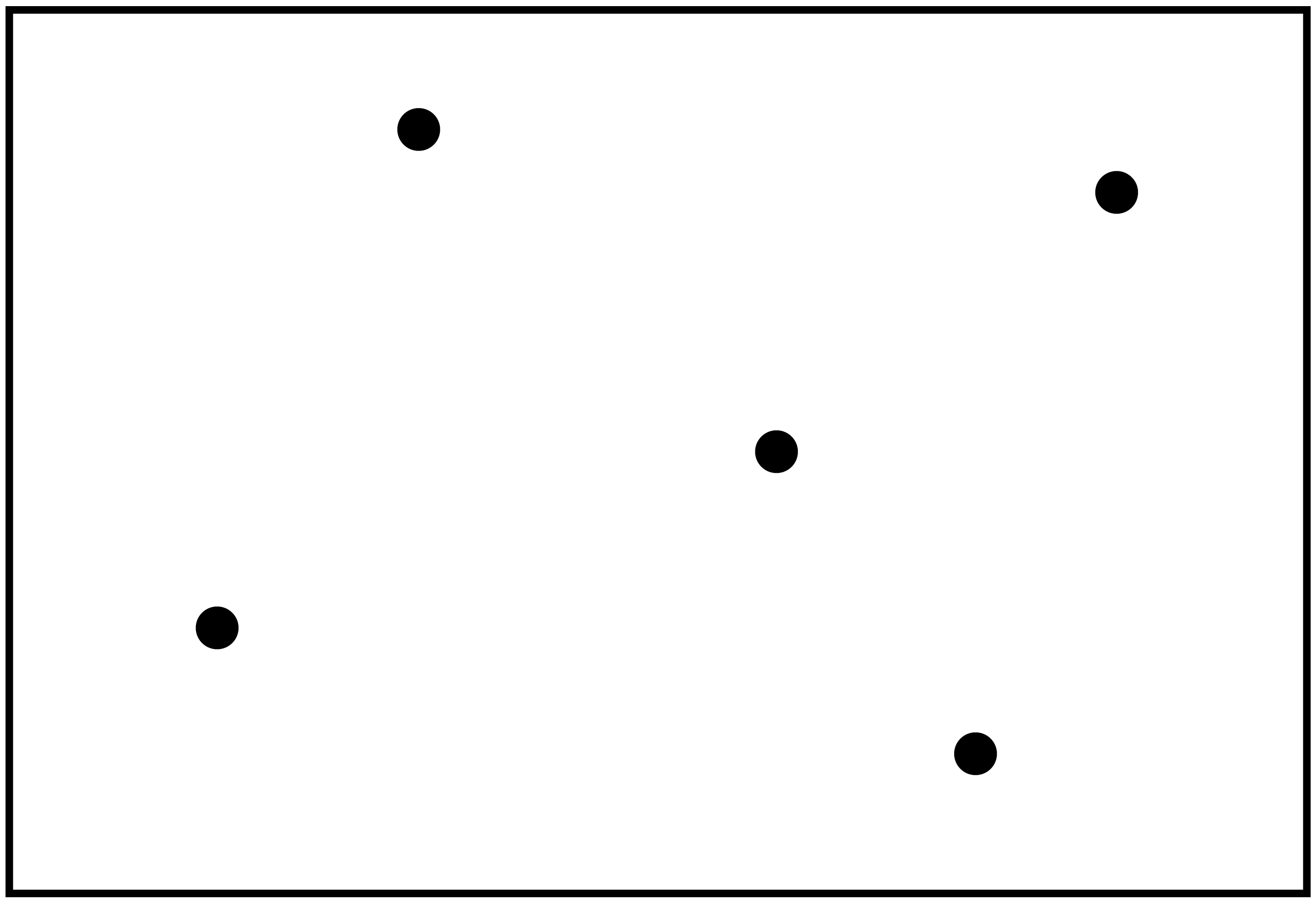
Random, countable set of points with:

- # points in set *A* is $\sim \mathrm{Poisson}[\nu(A)]$
- Independent #s of points in disjoint sets

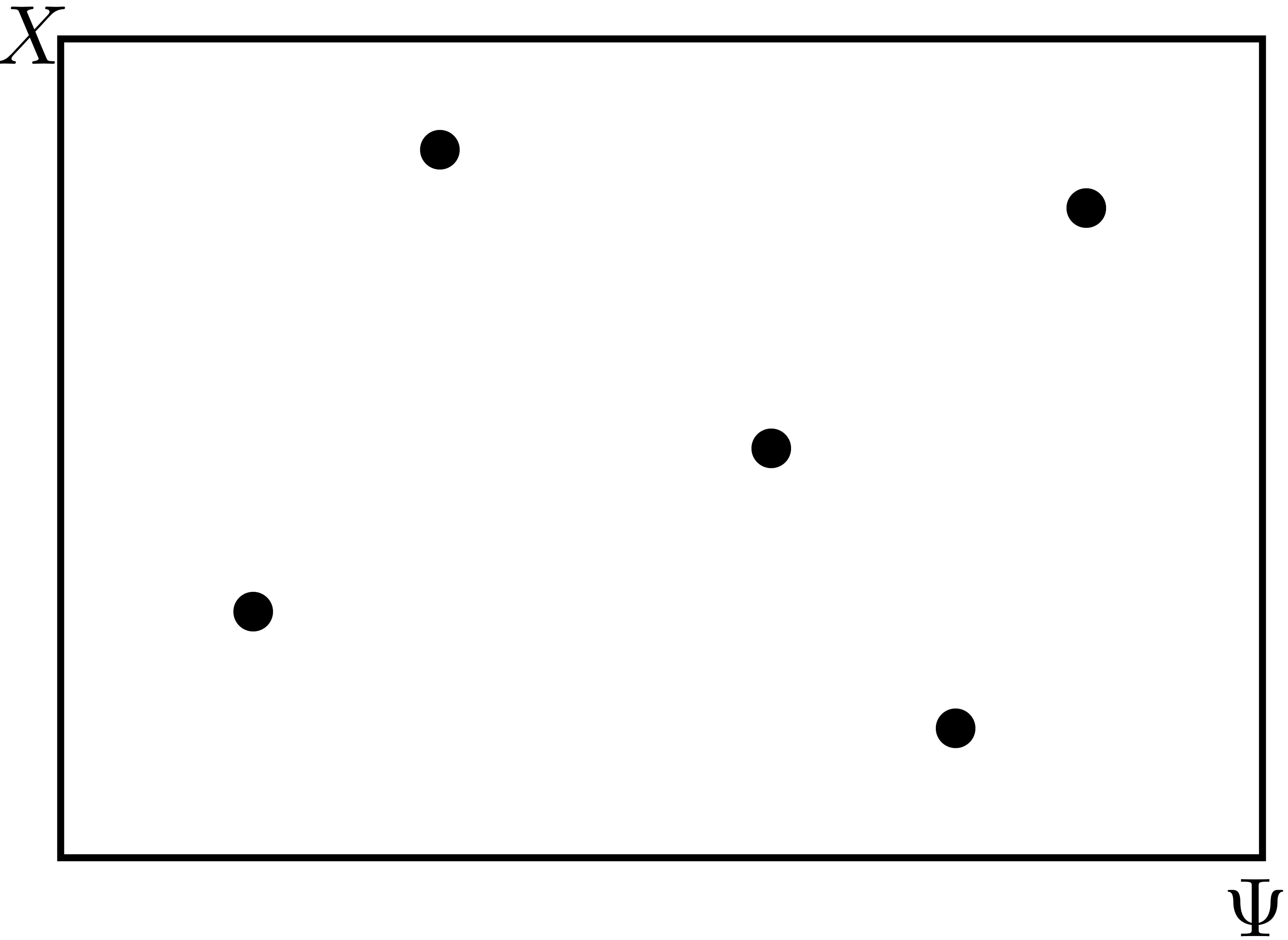


[Kingman 1993]

Poisson process fact: Thinning

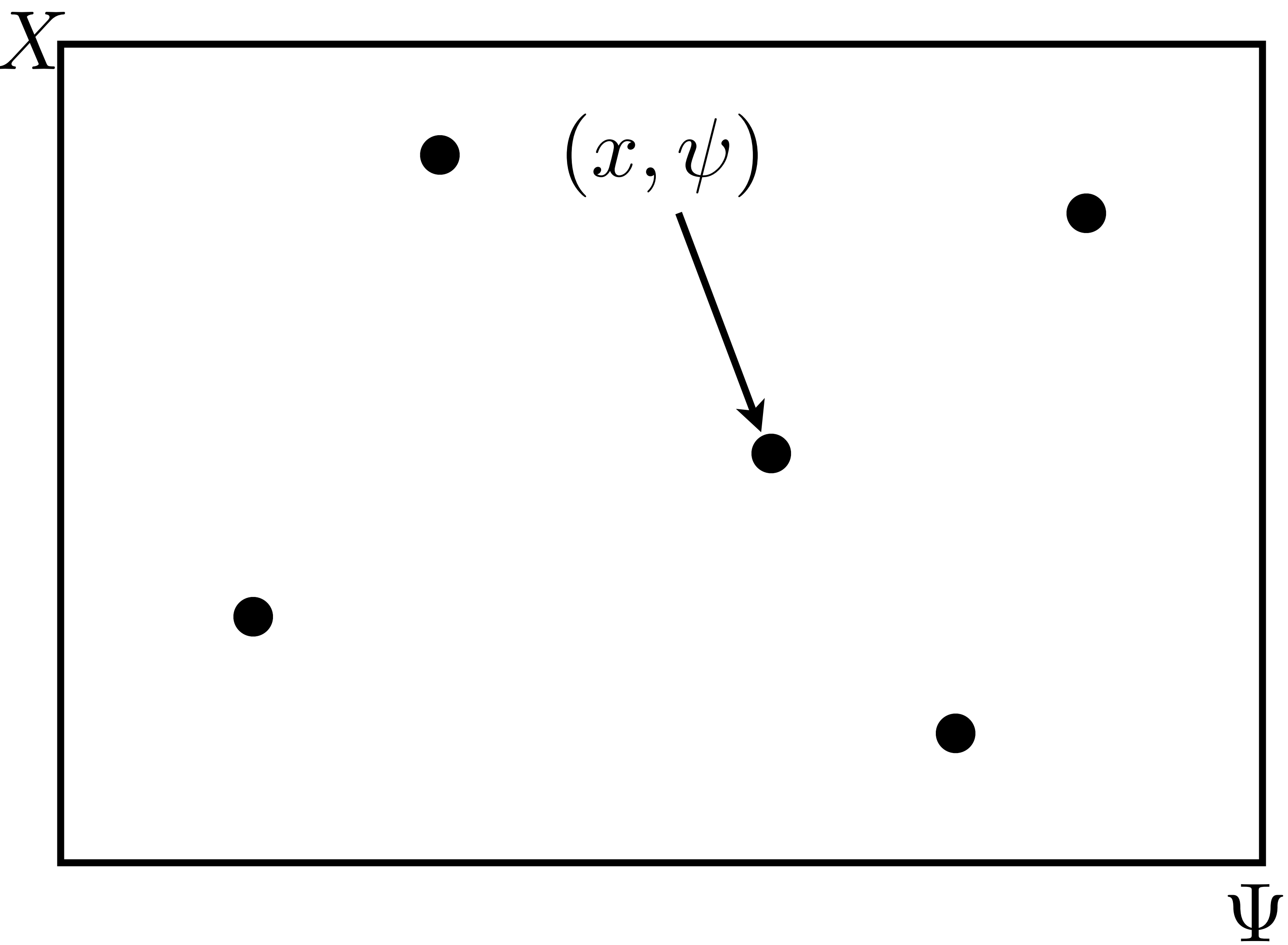


Poisson process fact: Thinning



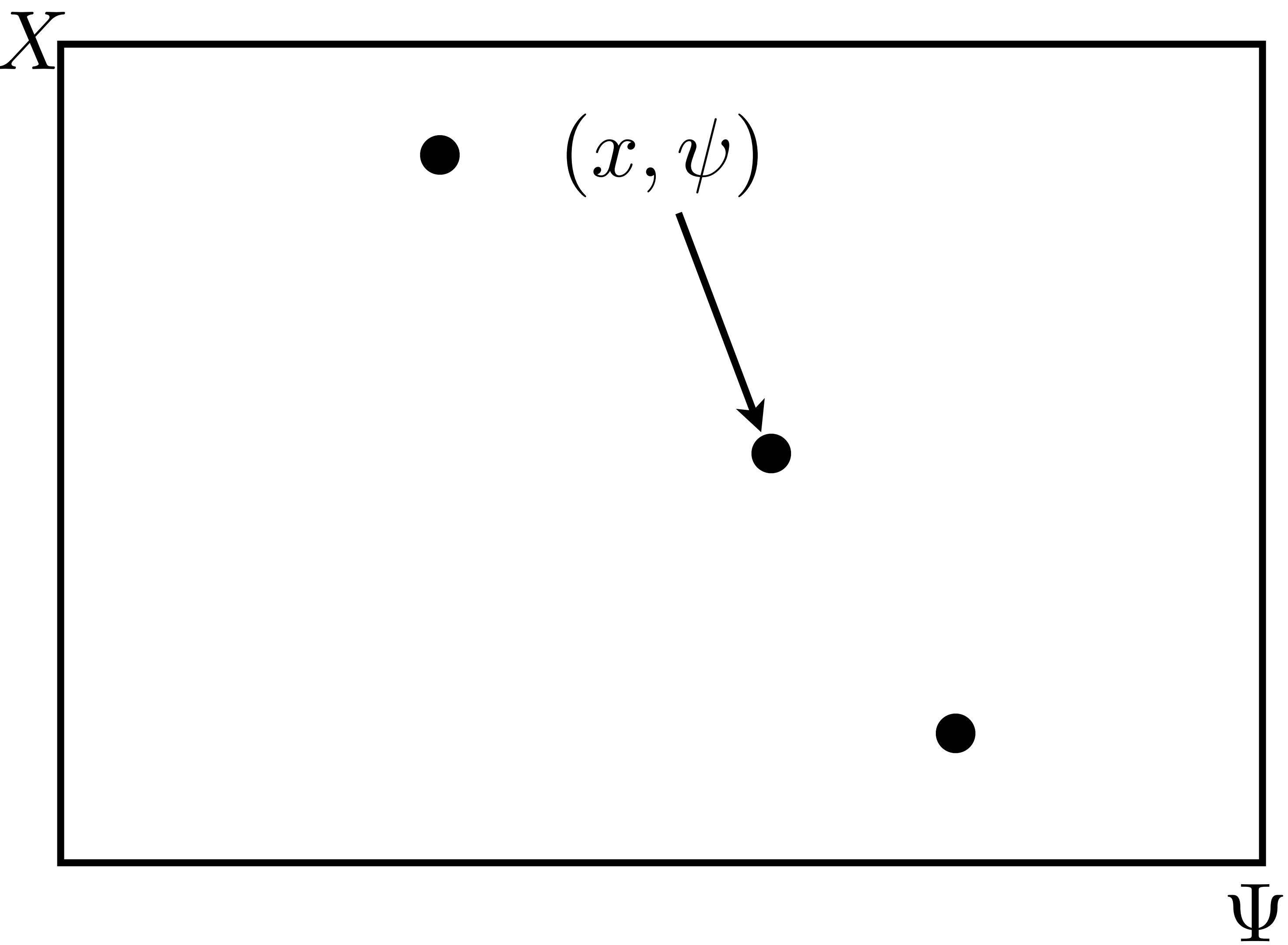
- Initial process has mean measure $\nu(dx, d\psi)$

Poisson process fact: Thinning



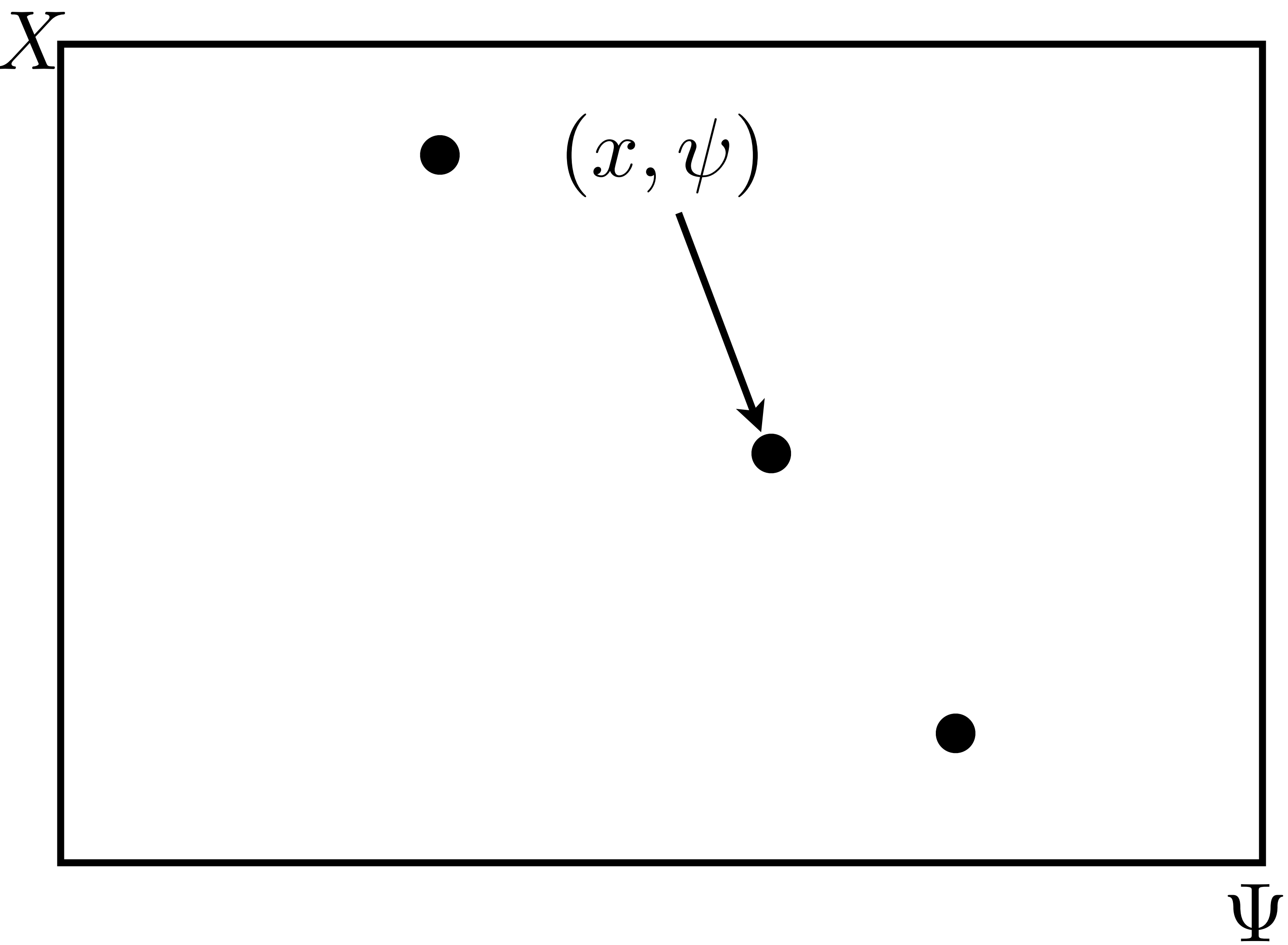
- Initial process has mean measure $\nu(dx, d\psi)$
- Keep point at (x, ψ) with probability $h(x, \psi)$

Poisson process fact: Thinning



- Initial process has mean measure $\nu(dx, d\psi)$
- Keep point at (x, ψ) with probability $h(x, \psi)$

Poisson process fact: Thinning



- Initial process has mean measure $\nu(dx, d\psi)$
- Keep point at (x, ψ) with probability $h(x, \psi)$
- Thinned process has mean measure

$$\nu_{thin}(A) = \int_{(x,\psi)\in A} \nu(dx, d\psi)\; h(x, \psi)$$

Beta process

Beta process

Two parts:

Beta process

Two parts:

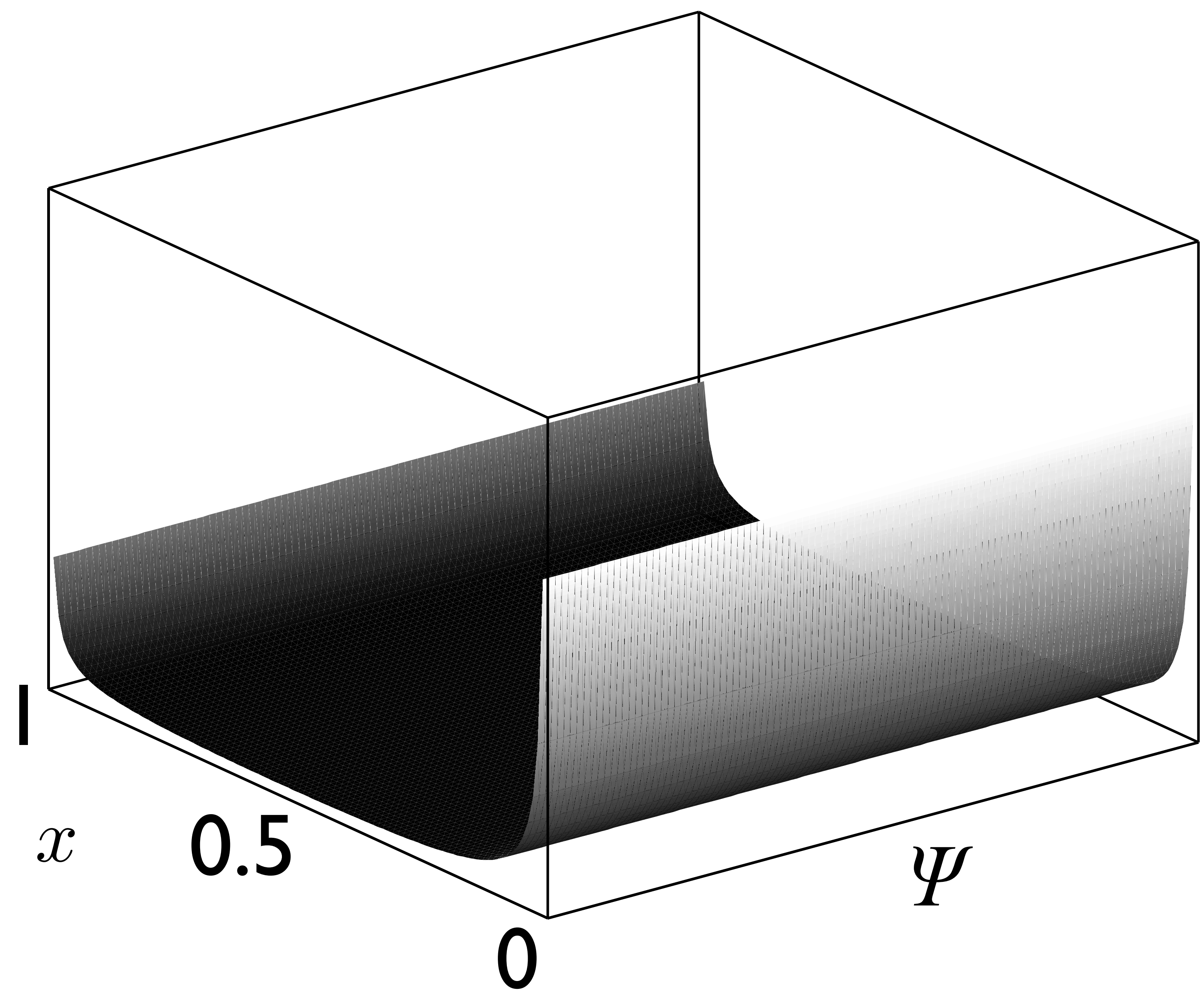
- Poisson process component

Beta process

Two parts:

- Poisson process component

$$\nu(dx, d\psi) = \gamma\theta x^{-1}(1-x)^{\theta-1}\, dx\, d\psi$$

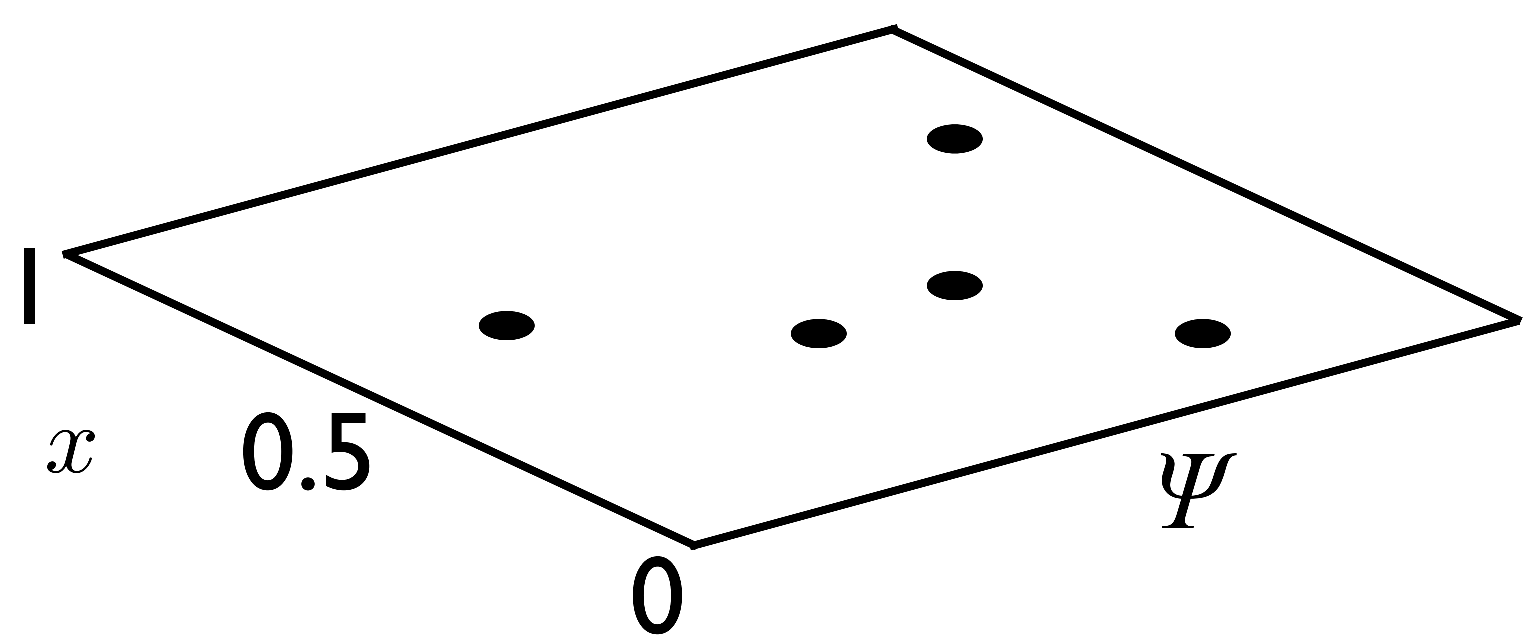


Beta process

Two parts:

- Poisson process component

$$\nu(dx, d\psi) = \gamma\theta x^{-1}(1-x)^{\theta-1}\, dx\, d\psi$$

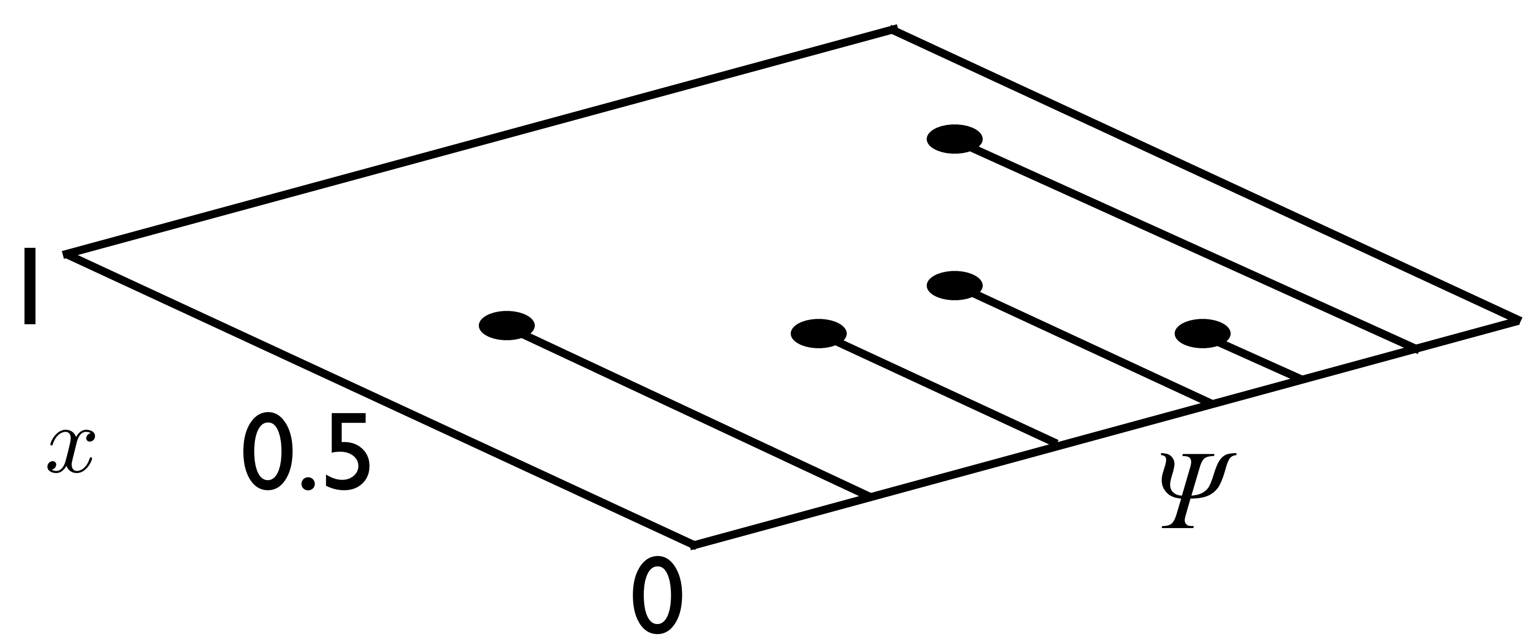


Beta process

Two parts:

- Poisson process component

$$\nu(dx, d\psi) = \gamma\theta x^{-1}(1-x)^{\theta-1}\, dx\, d\psi$$

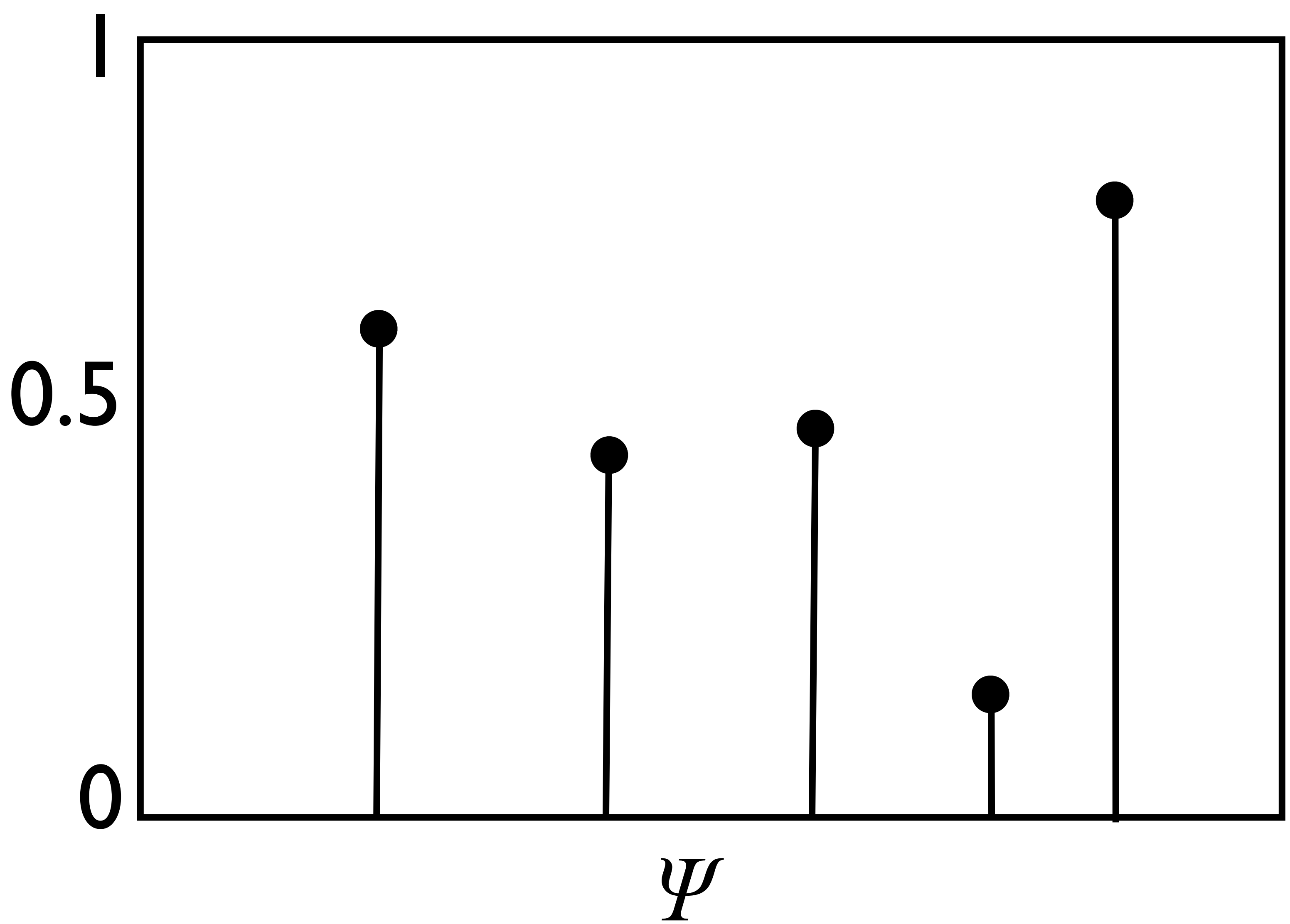


Beta process

Two parts:

- Poisson process component

$$\nu(dx, d\psi) = \gamma\theta x^{-1}(1-x)^{\theta-1}\, dx\, d\psi$$



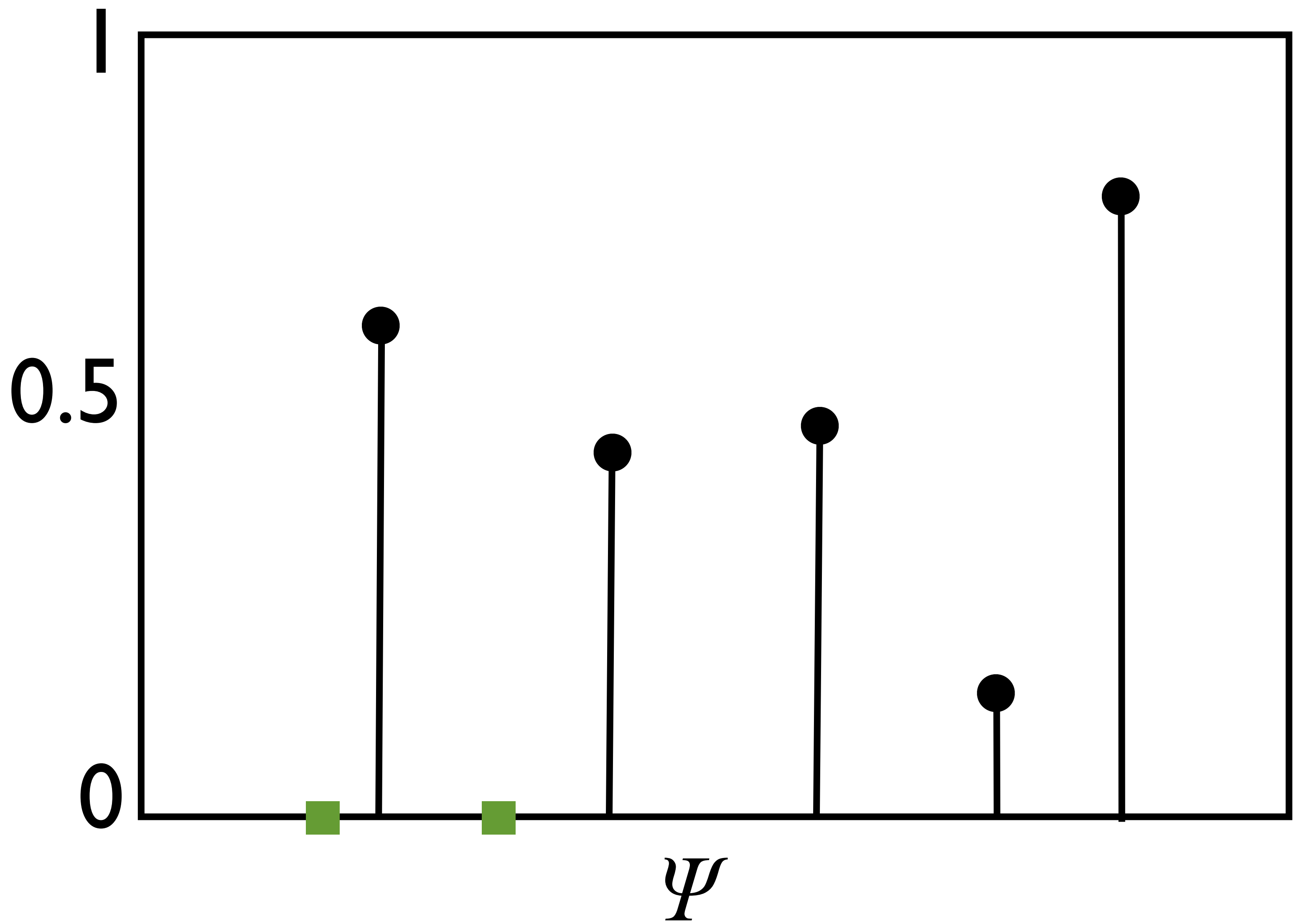
Beta process

Two parts:

- Poisson process component

$$\nu(dx, d\psi) = \gamma\theta x^{-1}(1-x)^{\theta-1}\, dx\, d\psi$$

- Fixed atoms



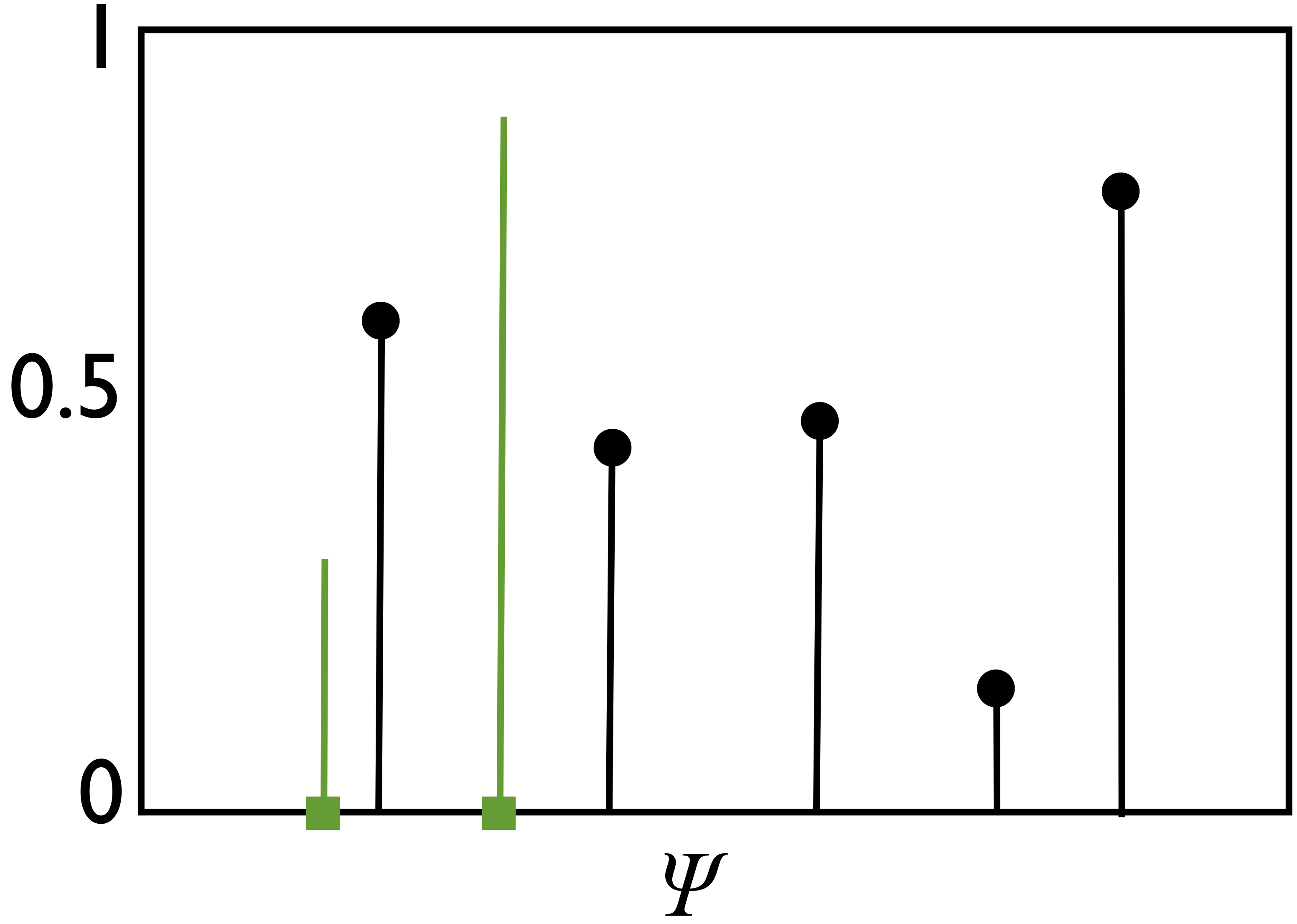
Beta process

Two parts:

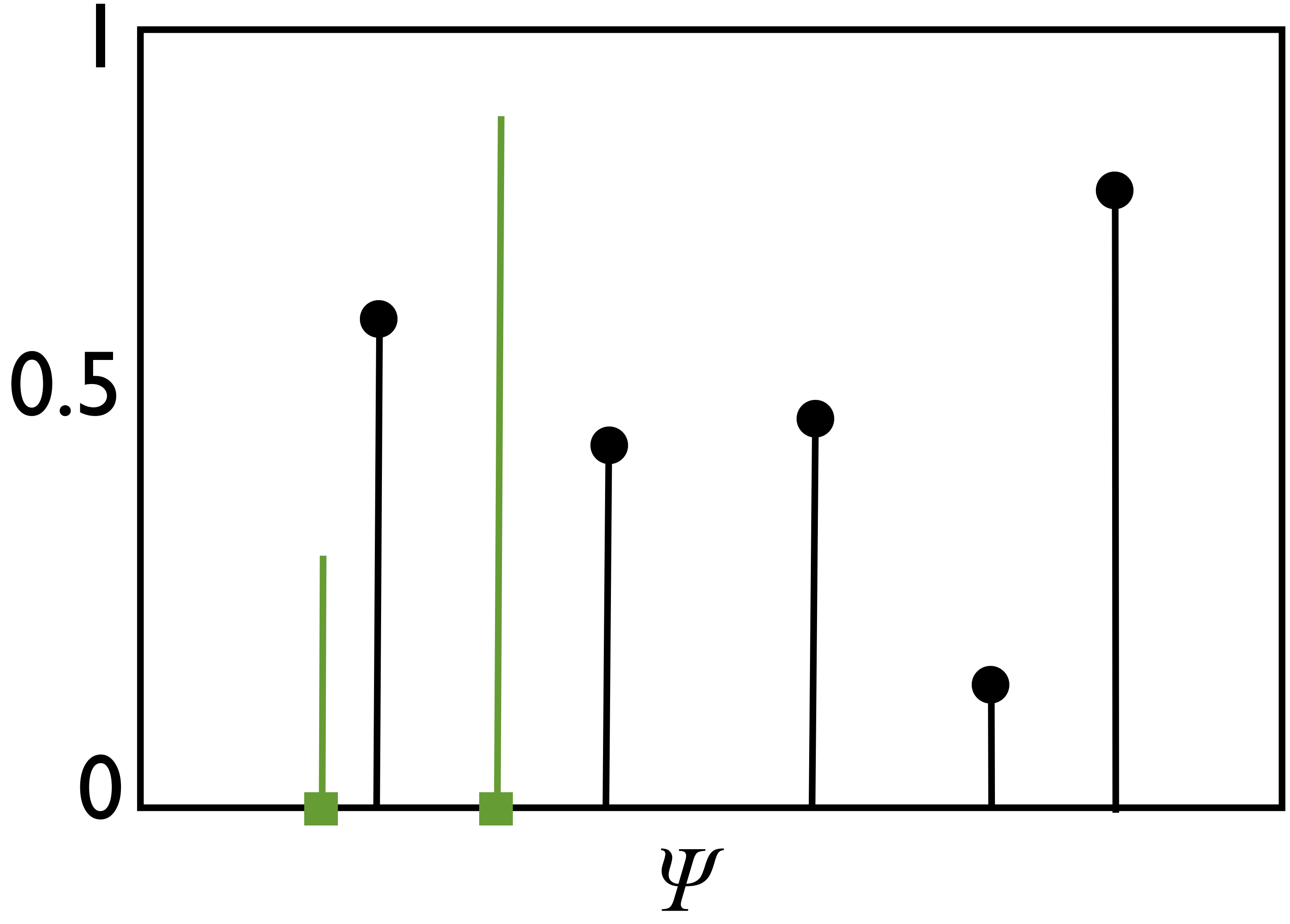
- Poisson process component

$$\nu(dx, d\psi) = \gamma\theta x^{-1}(1-x)^{\theta-1}\, dx\, d\psi$$

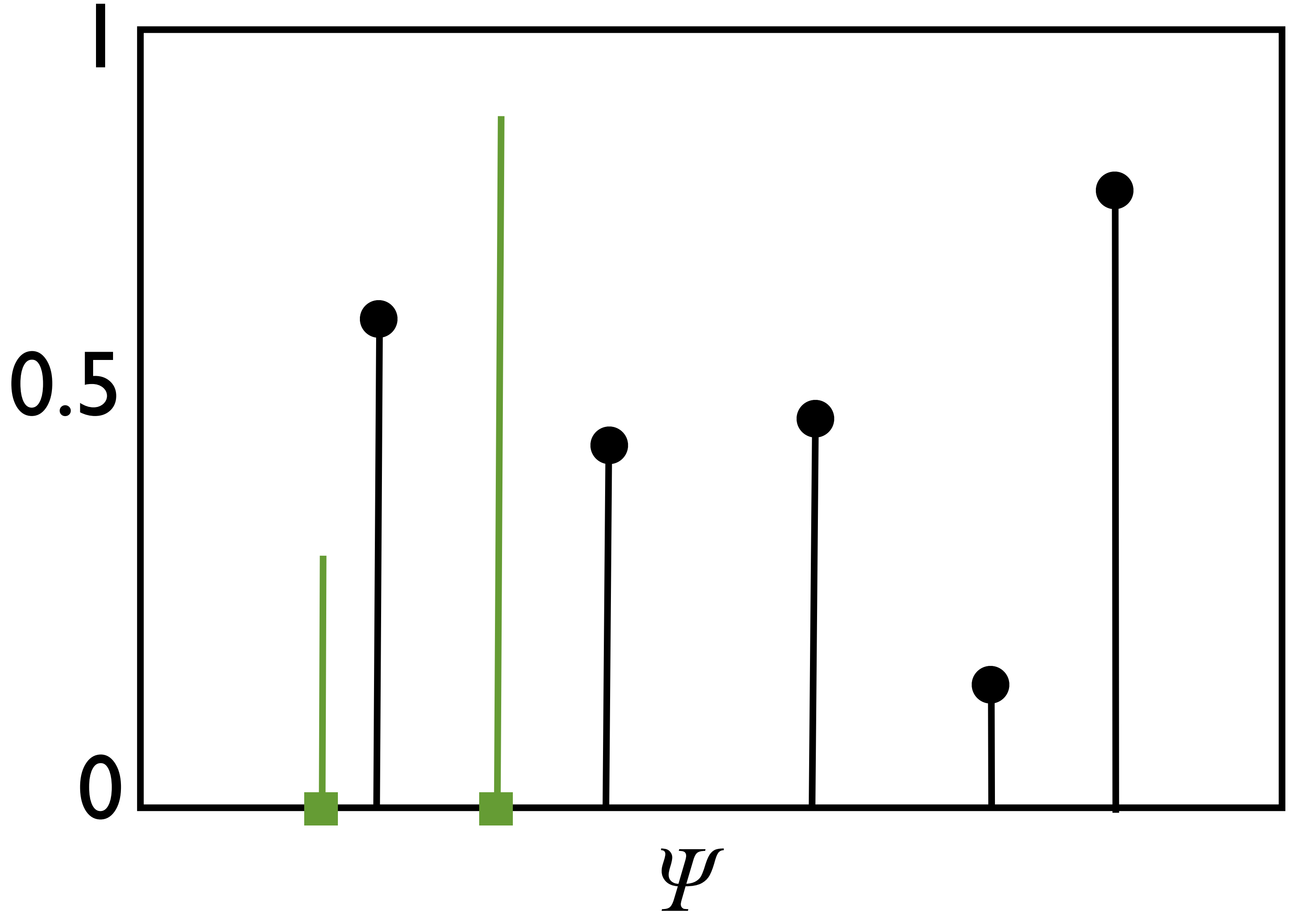
- Fixed atoms $x'_j \sim \mathrm{Beta}(a_j, b_j)$ at ψ'_j



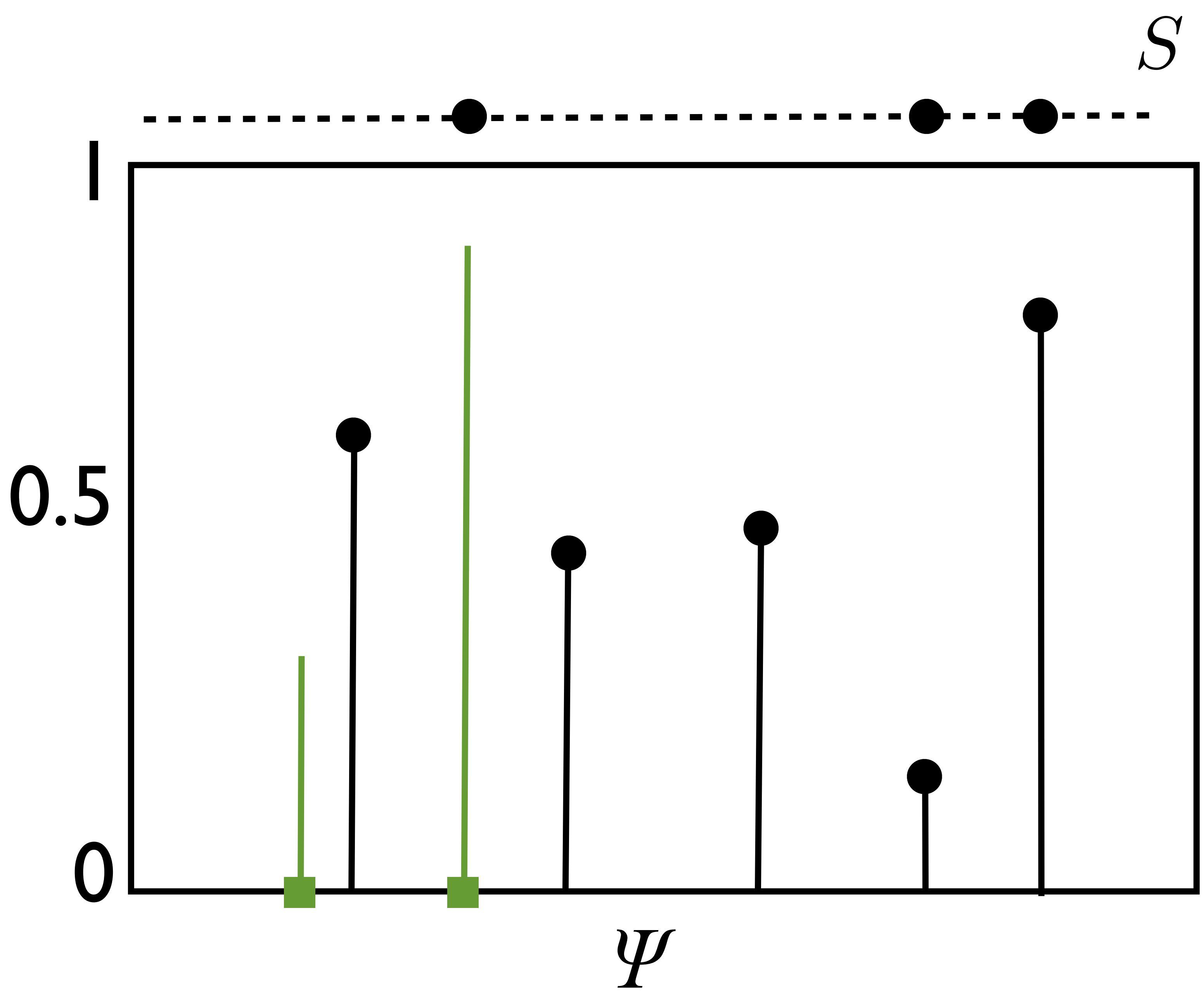
- Prior: beta process



- Prior: beta process
- Likelihood: Bernoulli process

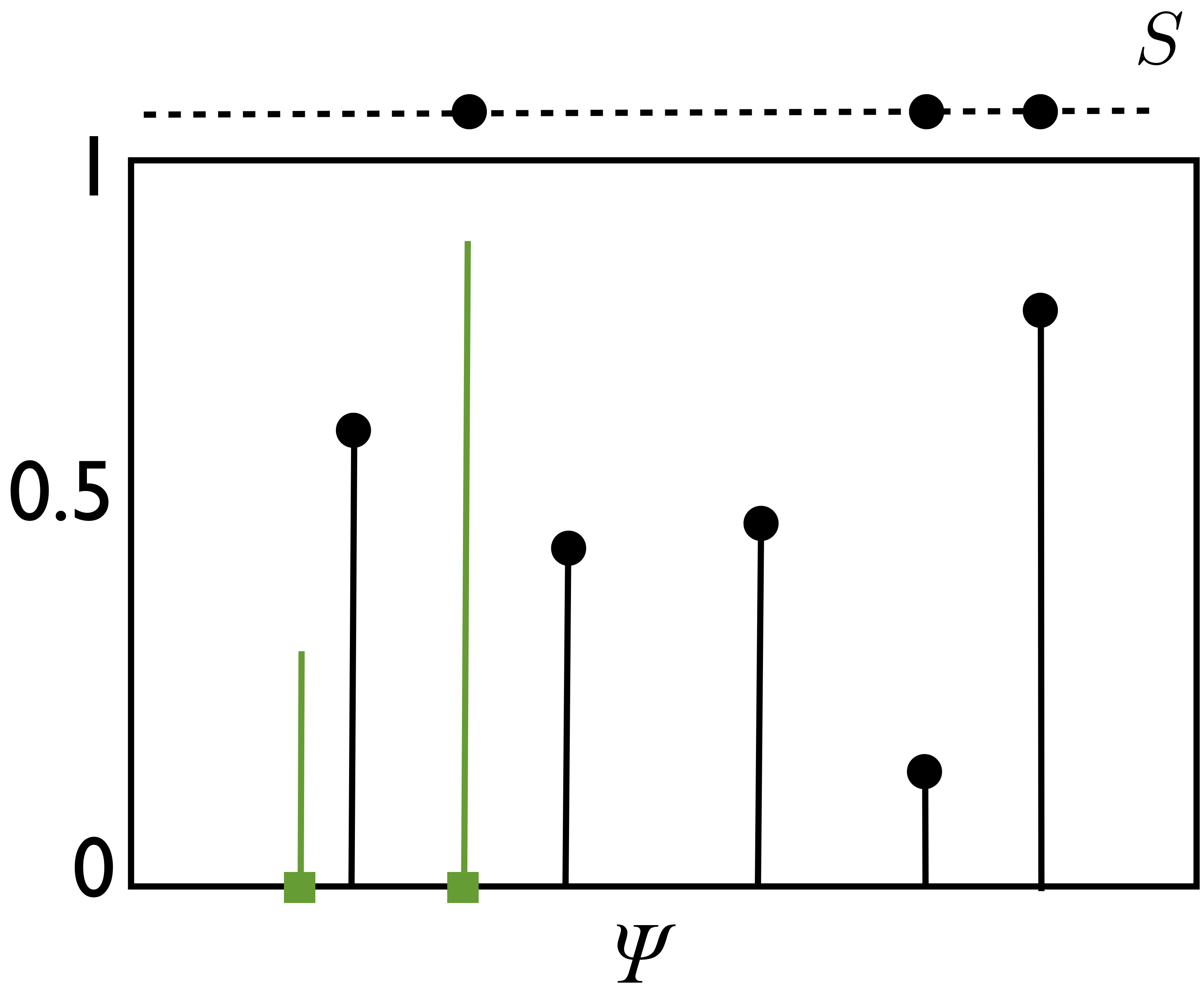


- Prior: beta process
- Likelihood: Bernoulli process



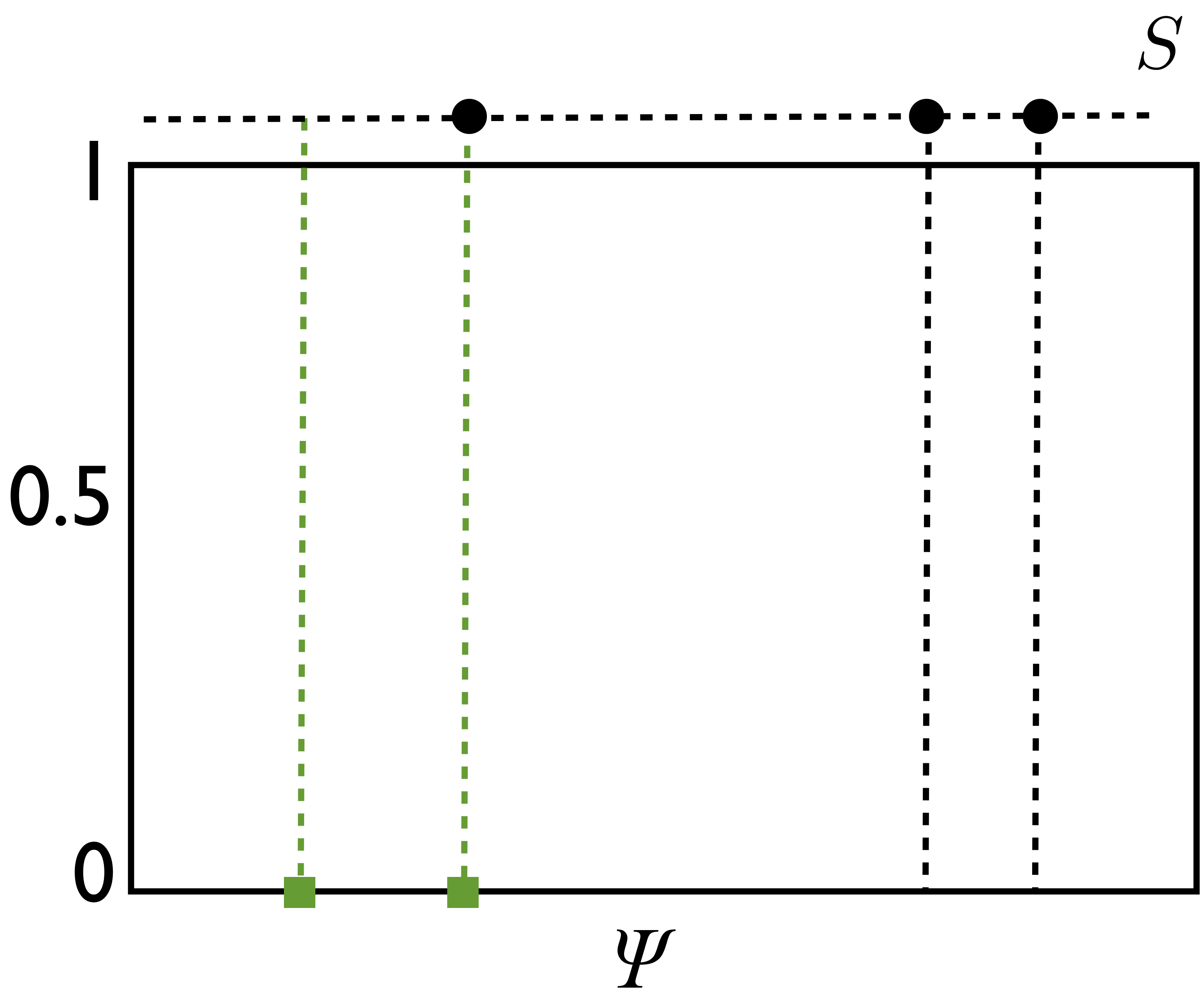
Posterior

- Prior: beta process
- Likelihood: Bernoulli process



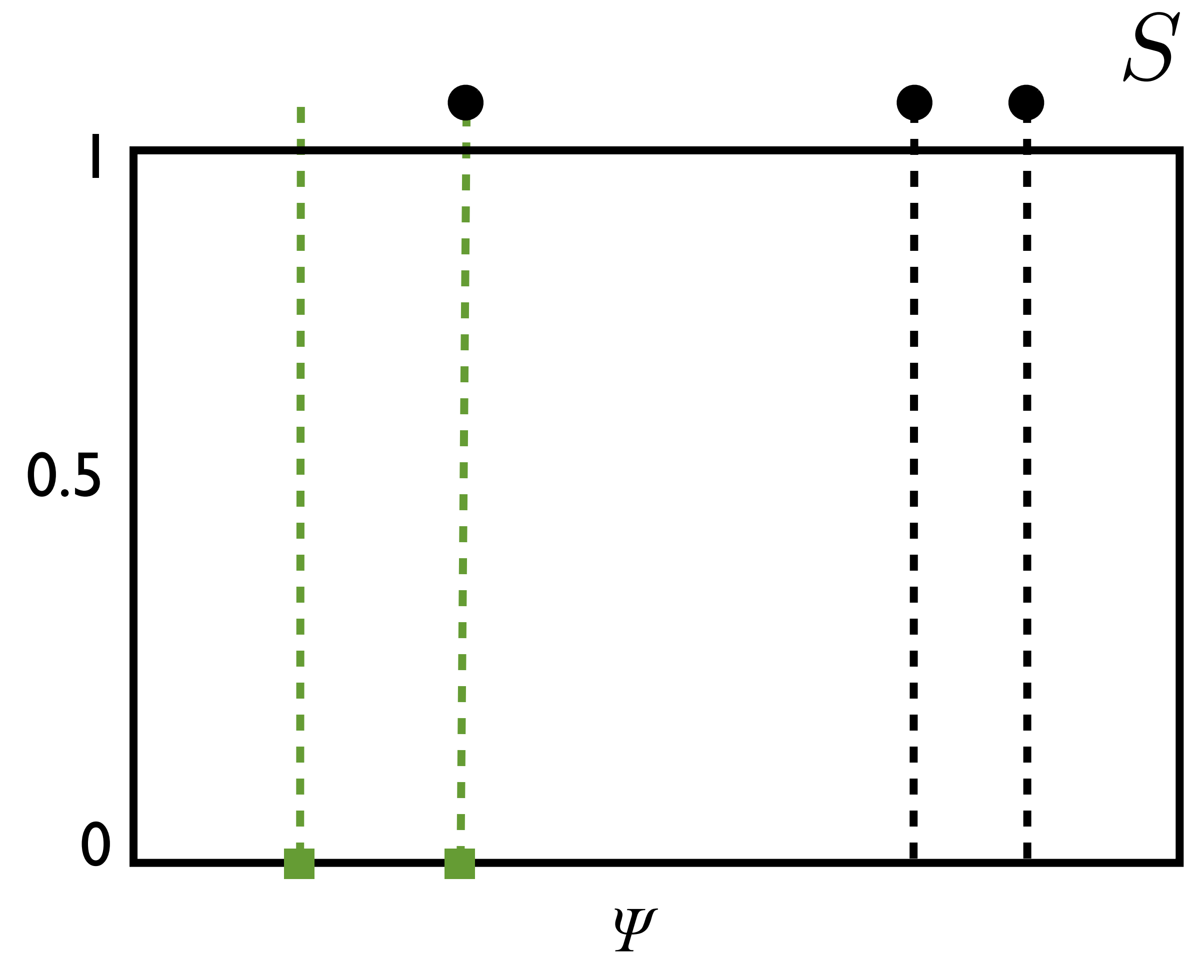
Posterior

- Prior: beta process
- Likelihood: Bernoulli process



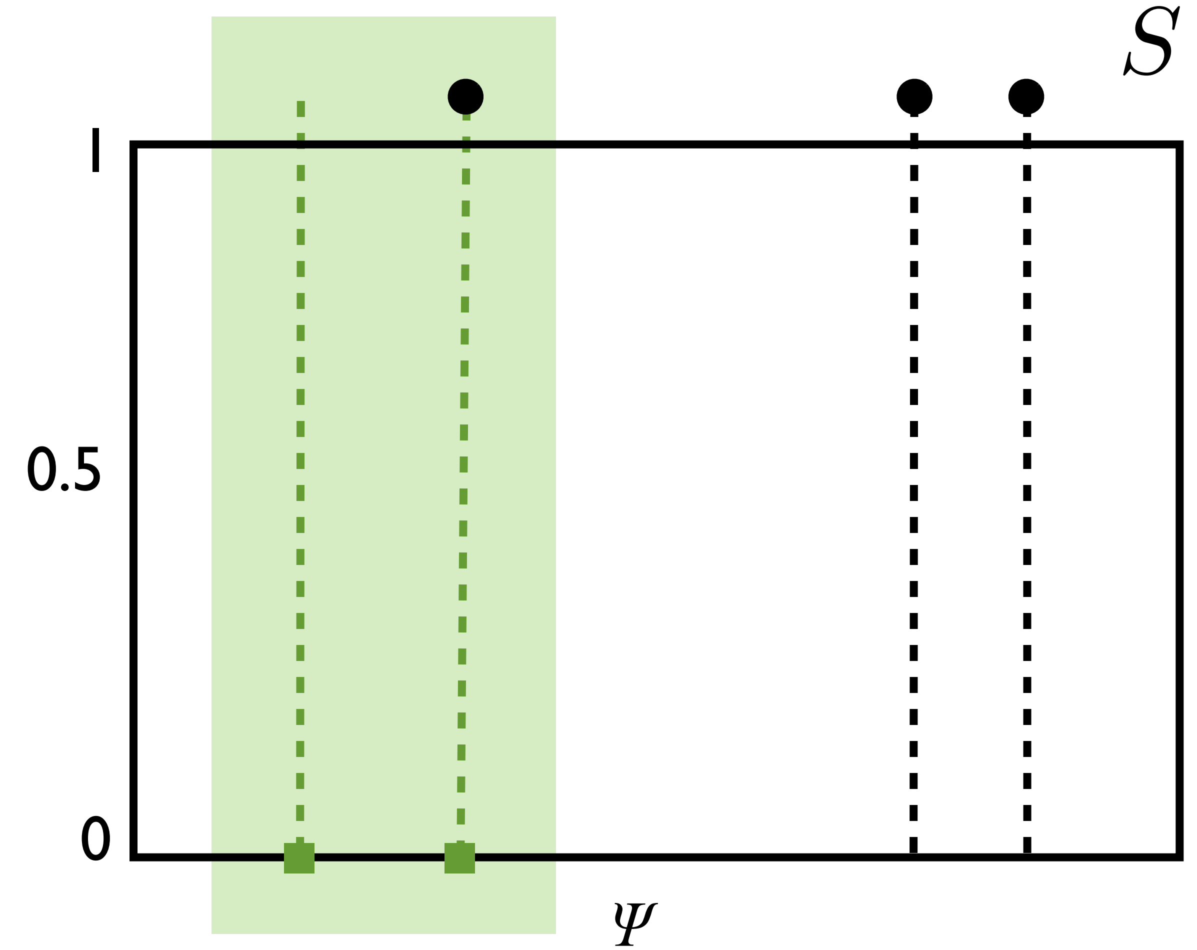
Posterior

- Prior: beta process
- Likelihood: Bernoulli process



Posterior

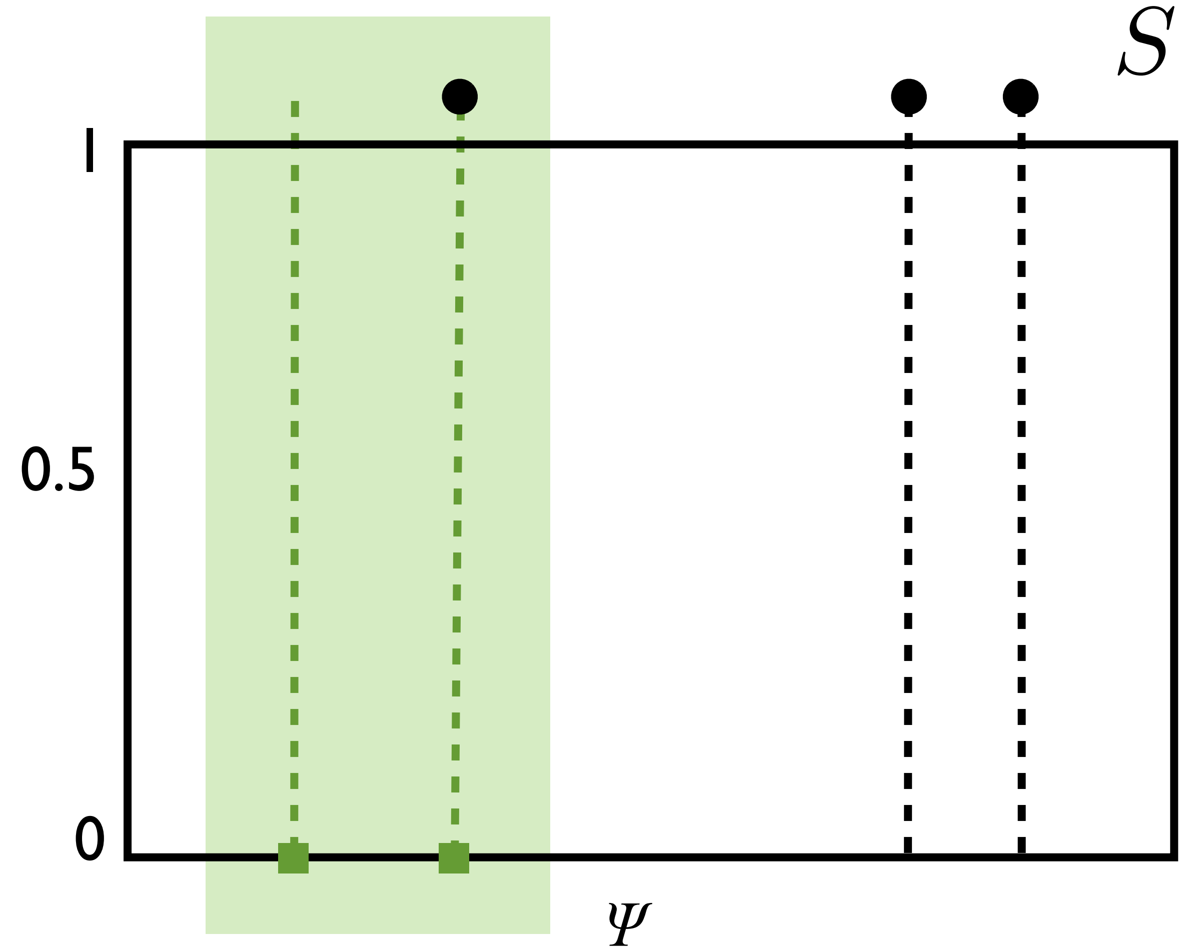
- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

Posterior

- Prior: beta process
- Likelihood: Bernoulli process

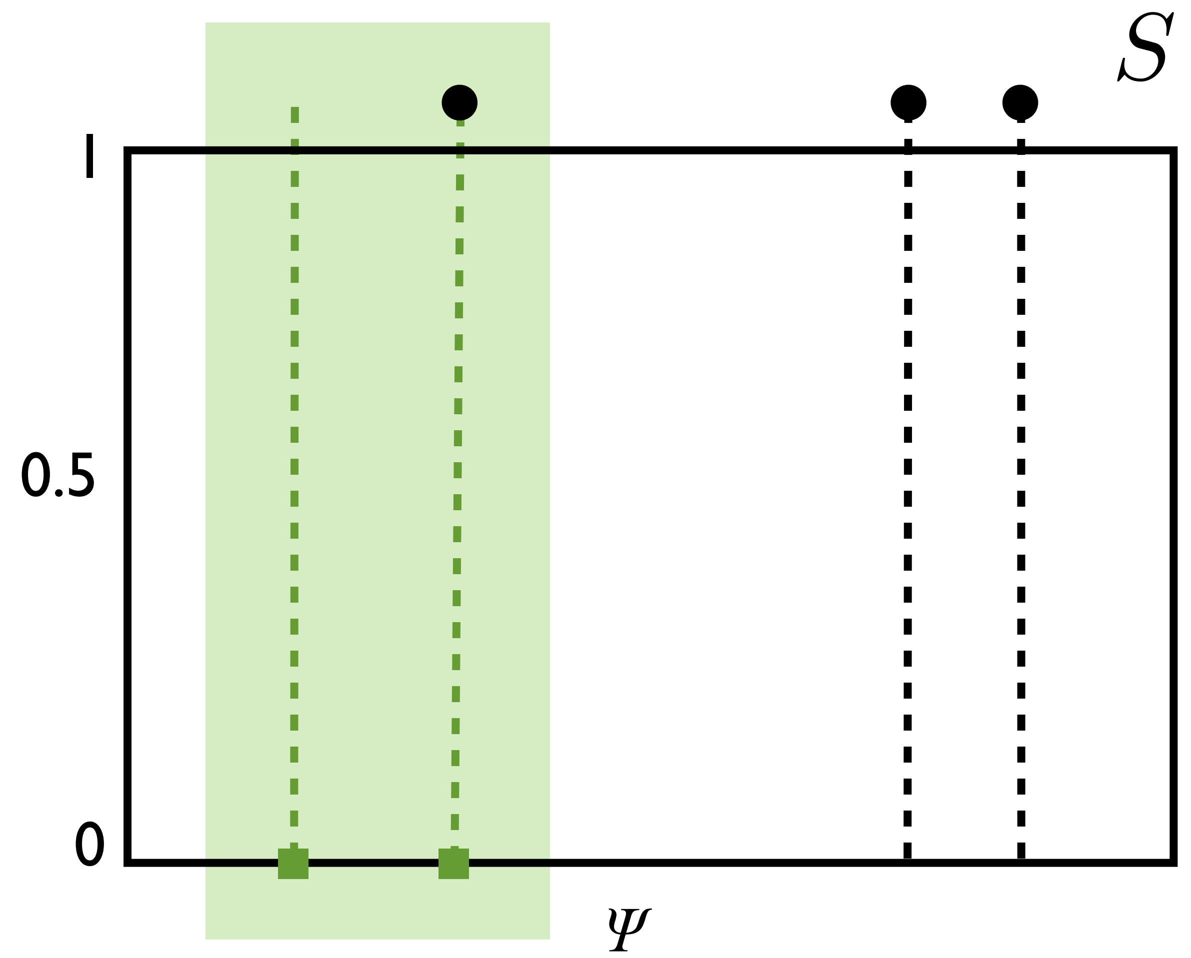


- Old fixed atom at ψ_j

$\mathrm{Beta}(x_j | a_j, b_j)$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process

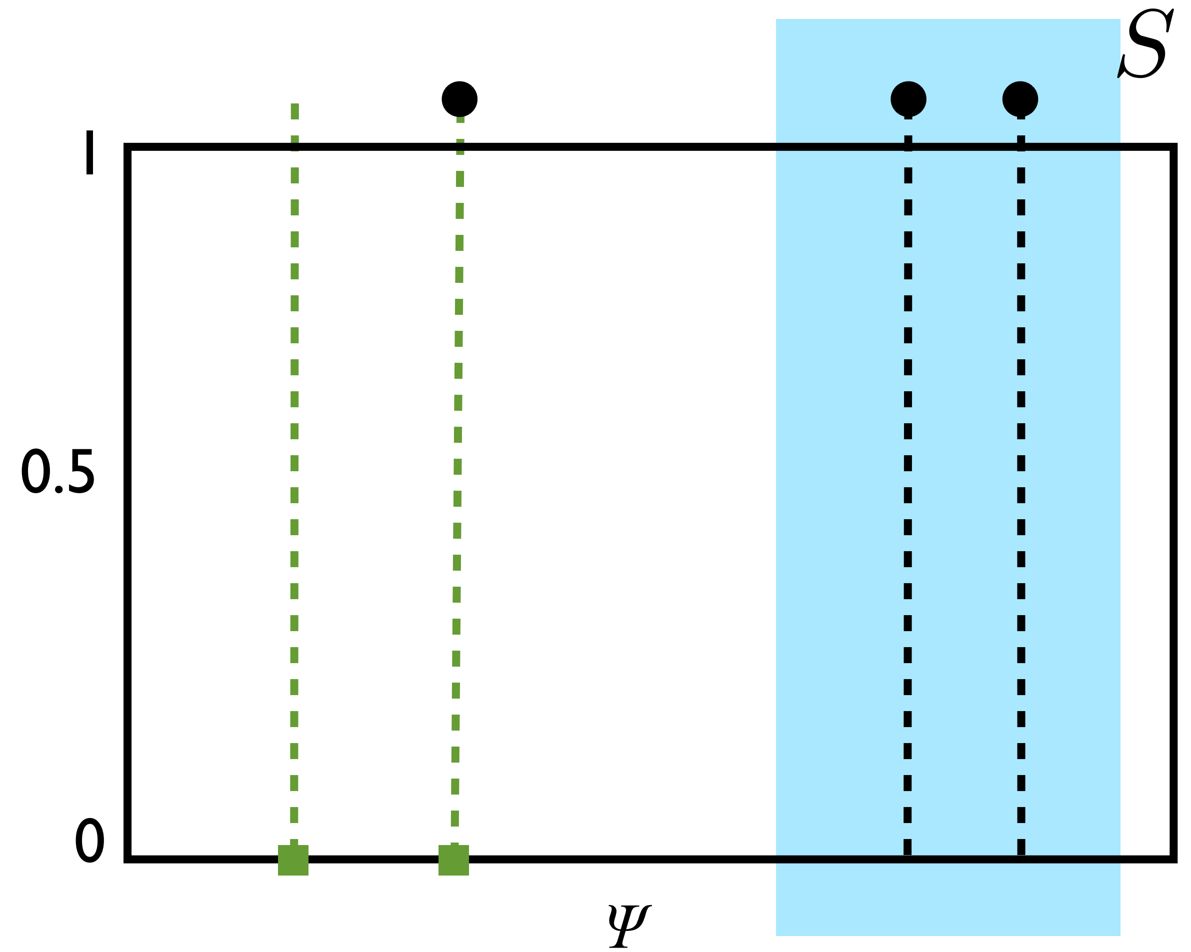


- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



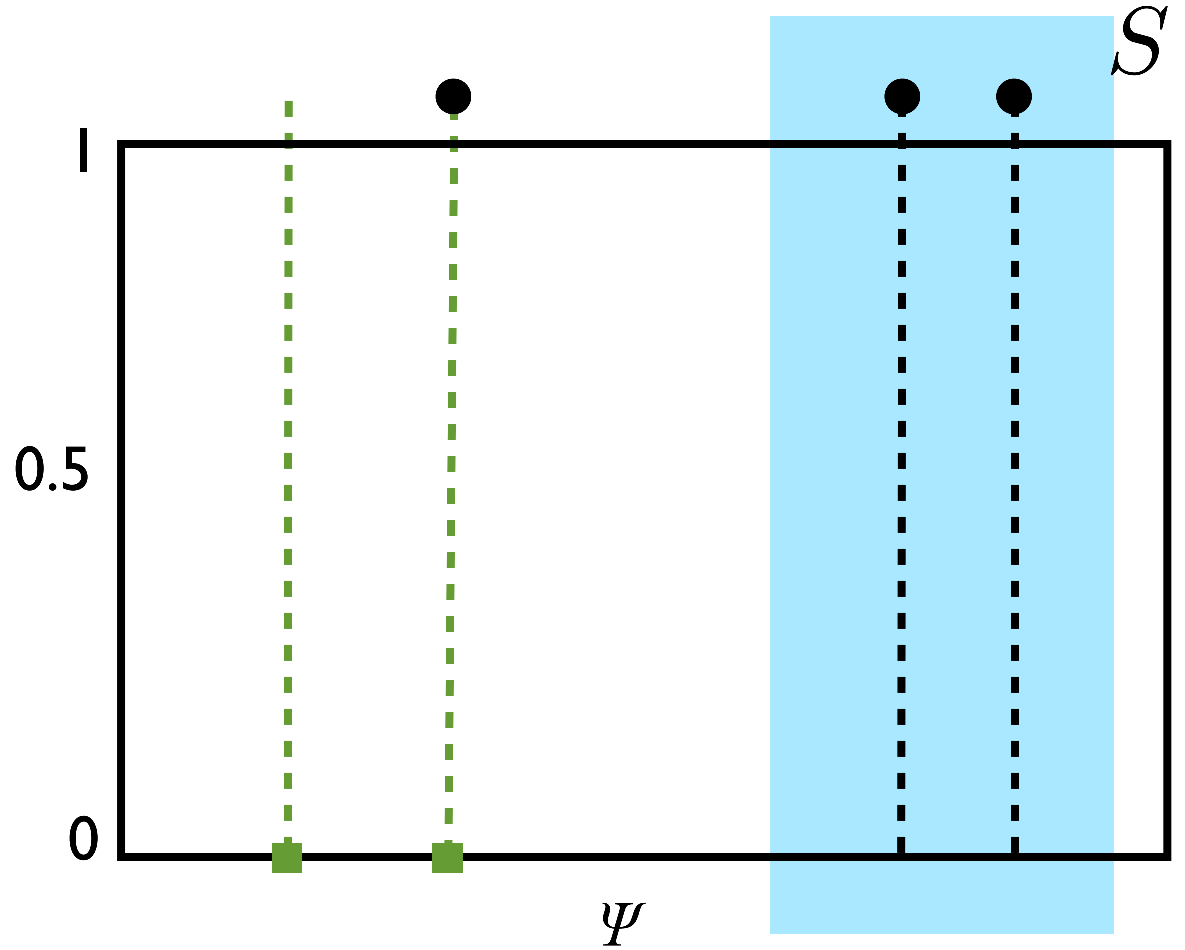
- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

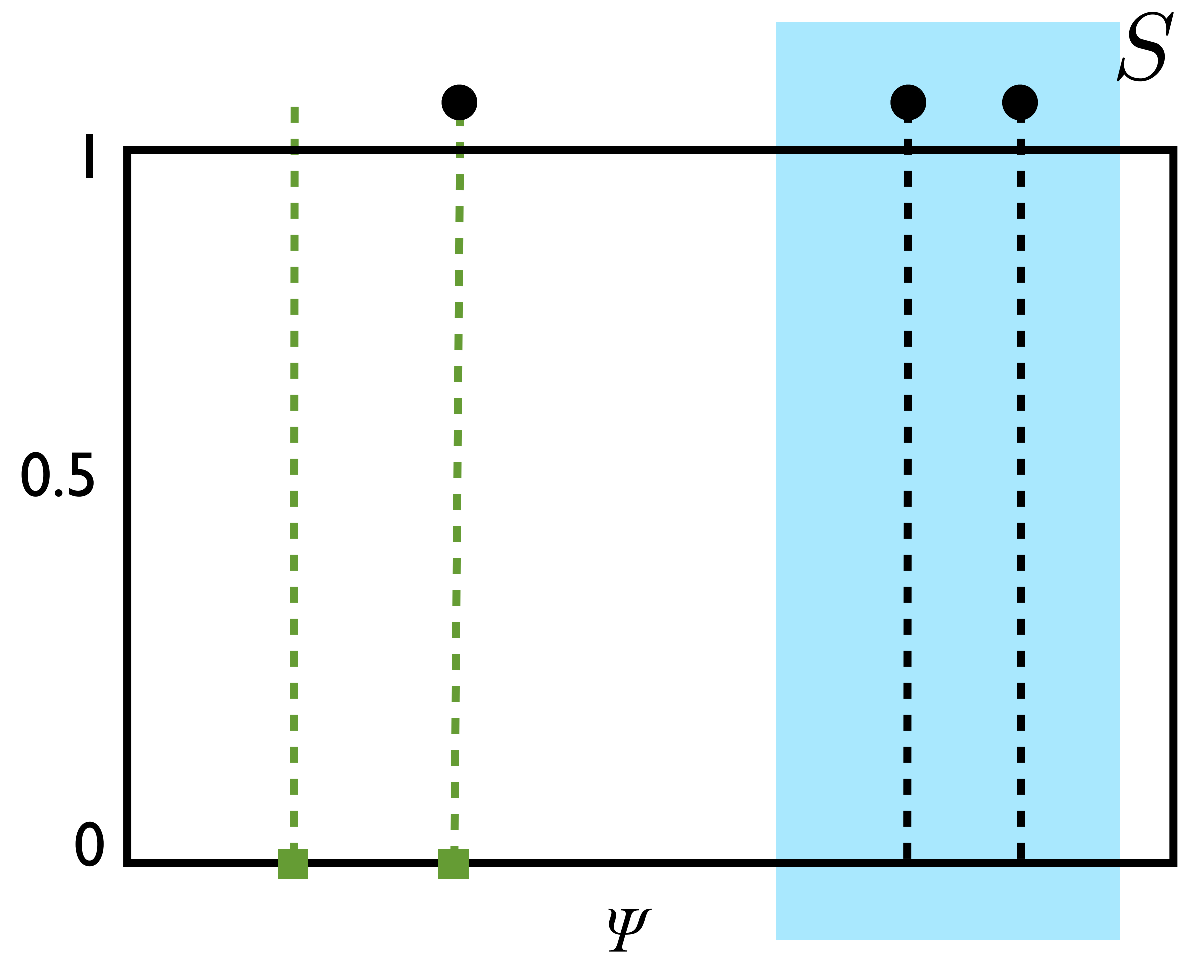
$$\mathrm{Beta}(x_j | a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j | a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

$$\gamma \theta x^{-1} (1-x)^{\theta - 1} \, dx$$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

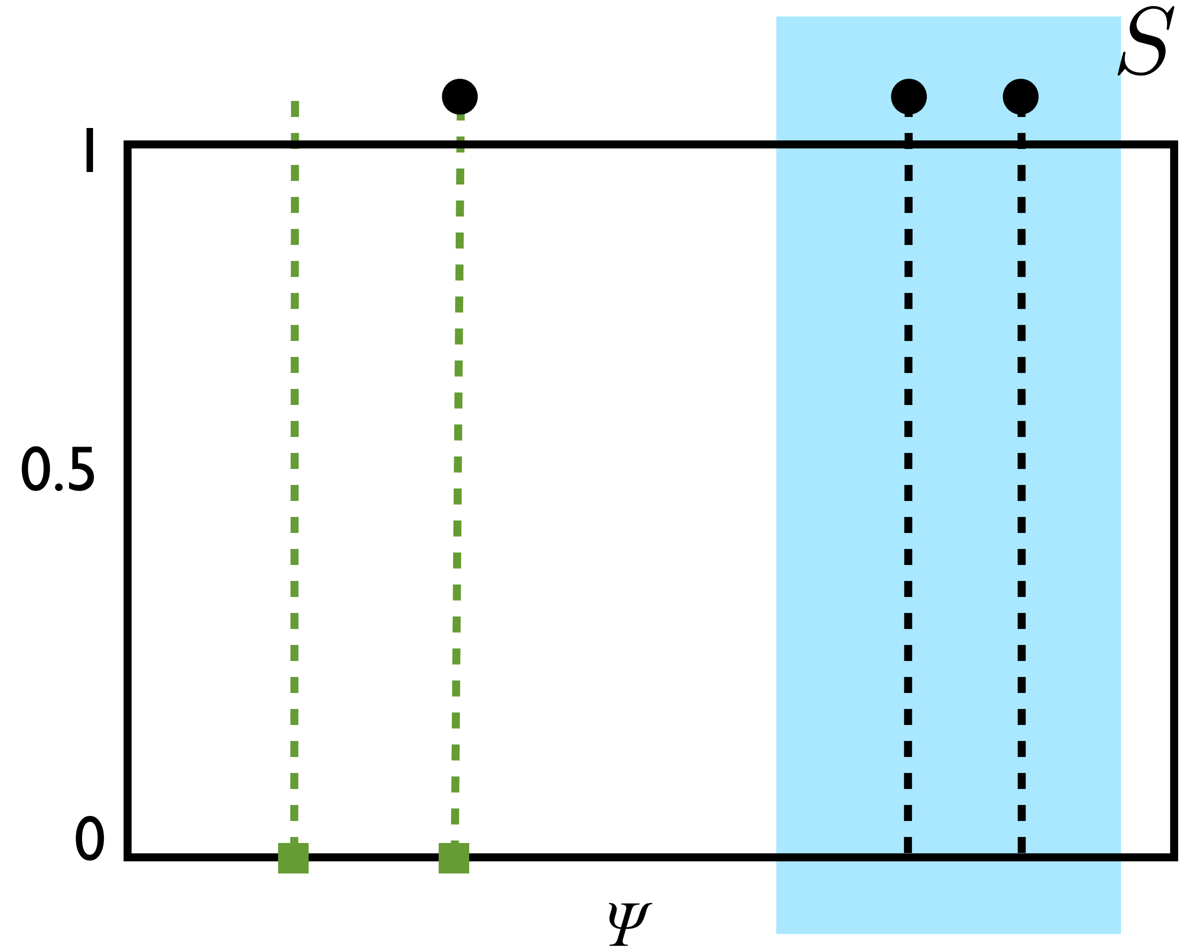
$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx \Rightarrow \qquad \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1}dx$$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

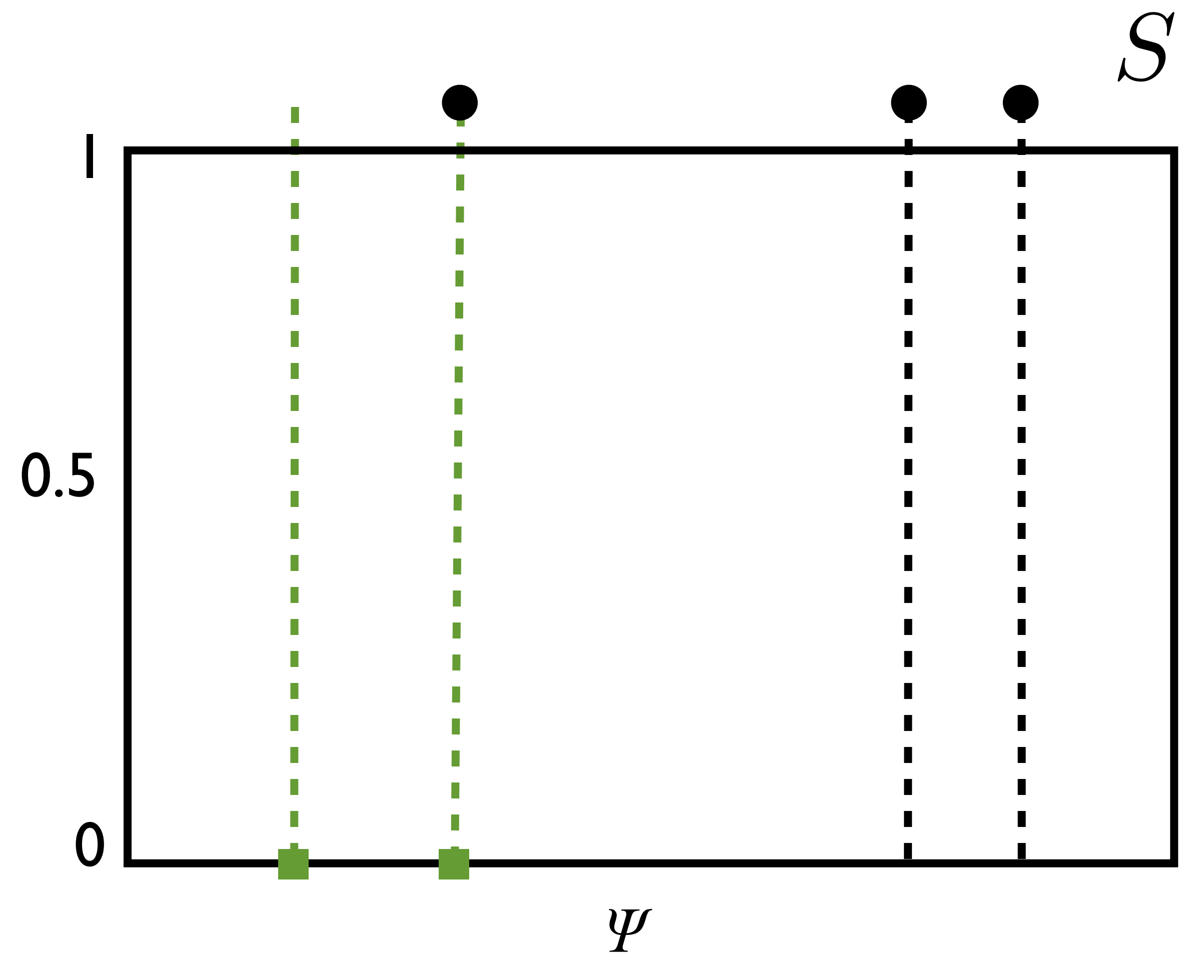
$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx \Rightarrow \mathrm{Beta}(x|1,\theta) \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1}dx$$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

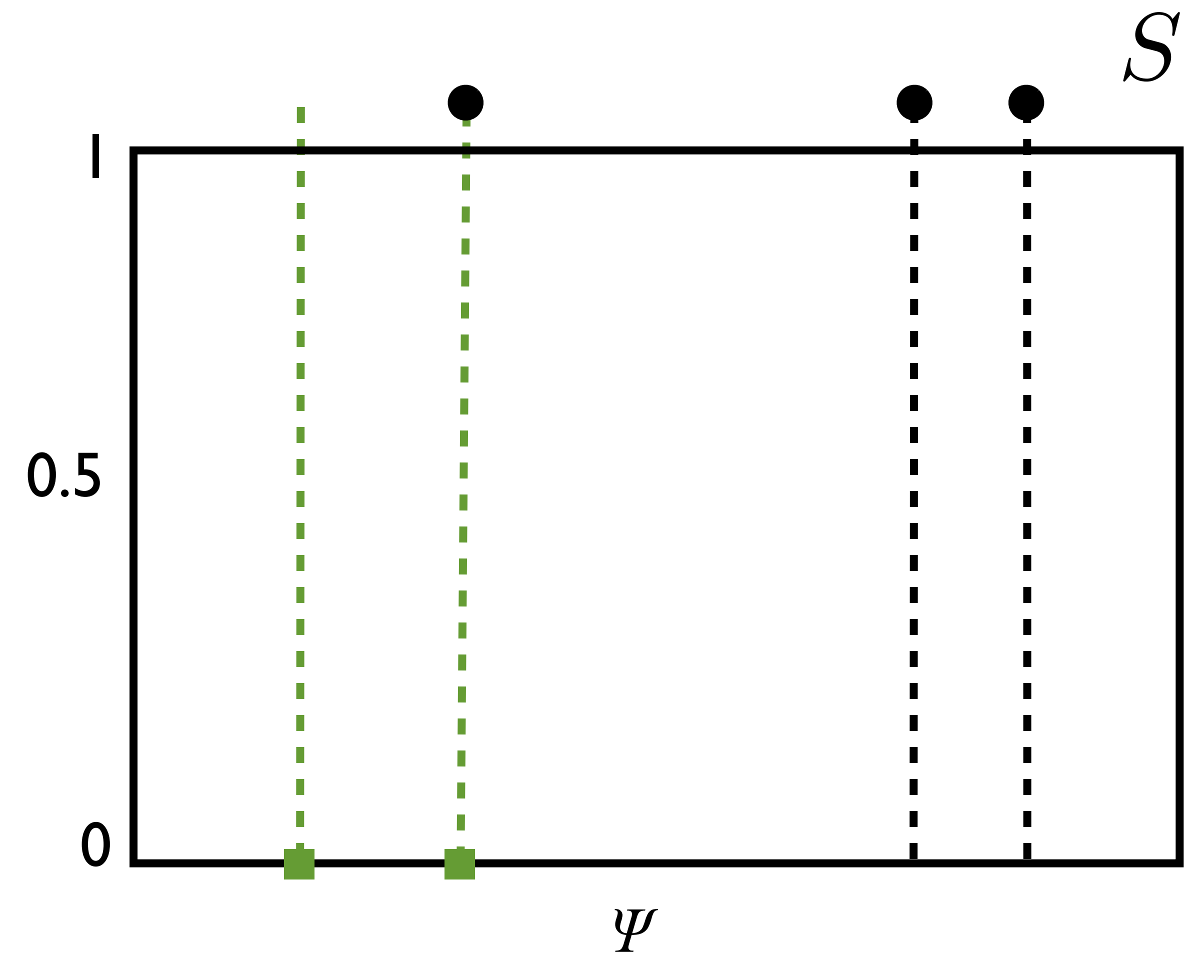
- New fixed atom at ψ

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx \Rightarrow \mathrm{Beta}(x|1,\theta) \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1} dx$$

- Unobserved Poisson process

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

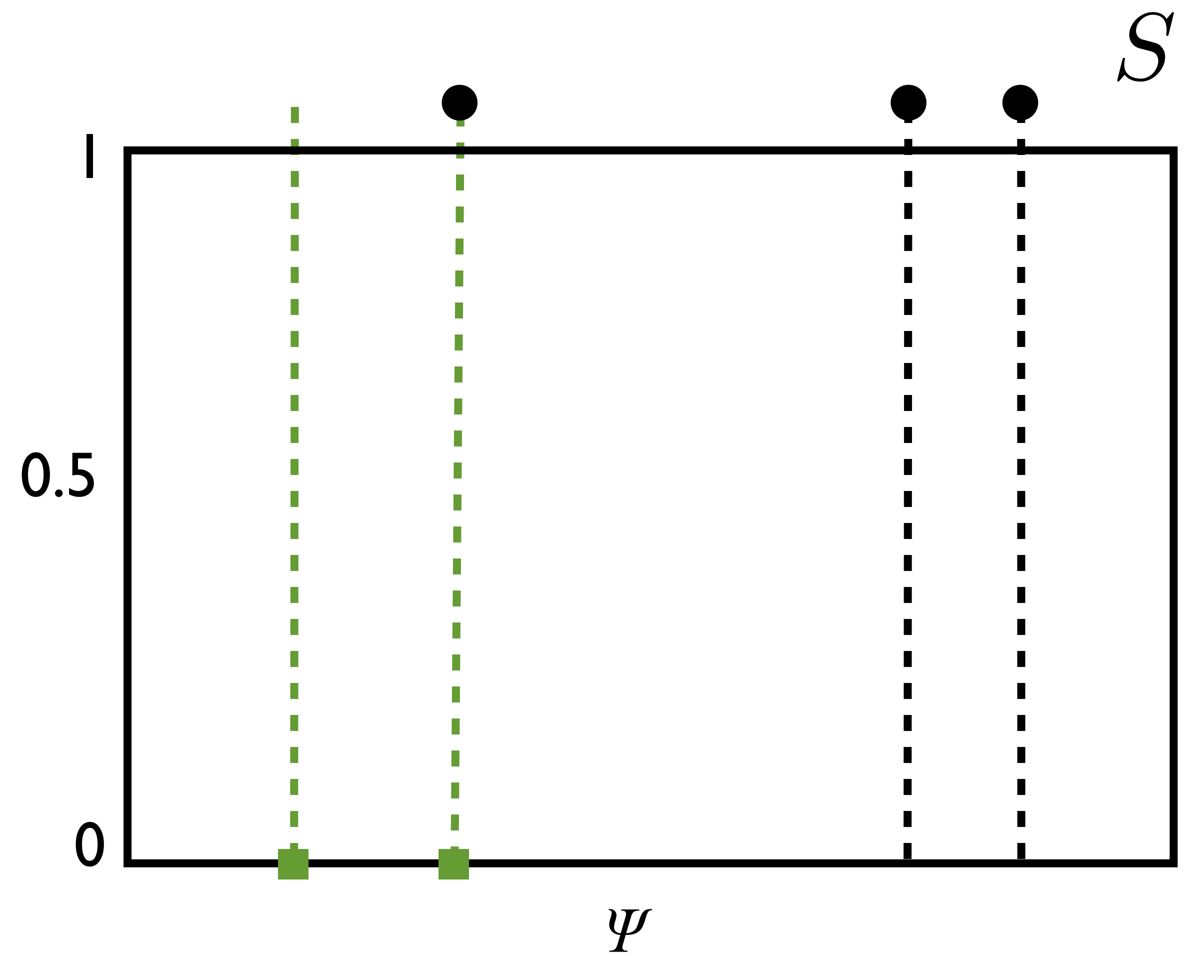
$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx \Rightarrow \mathrm{Beta}(x|1,\theta) \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1}dx$$

- Unobserved Poisson process

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx$$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

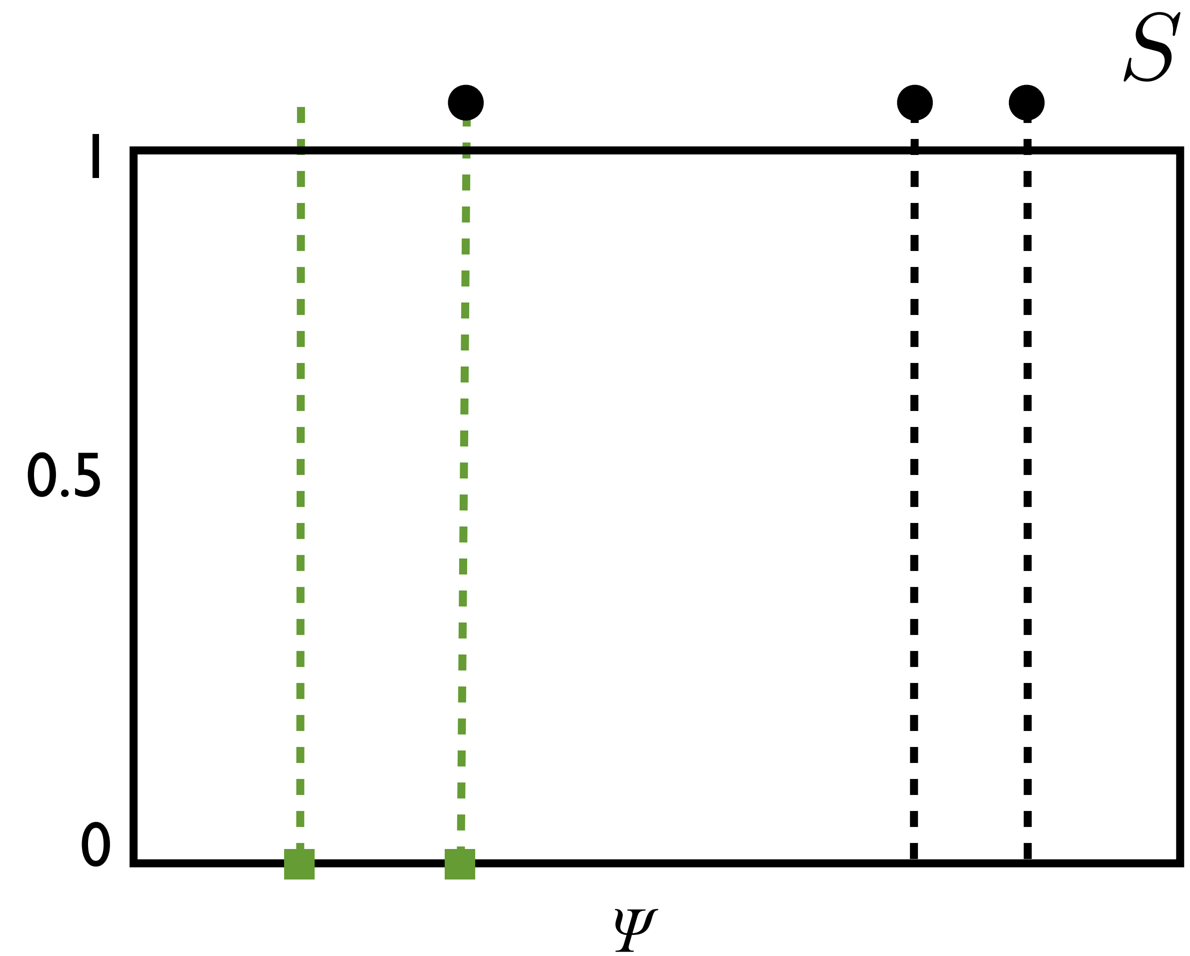
$$\gamma\theta x^{-1}(1-x)^{\theta-1}\, dx \Rightarrow \mathrm{Beta}(x|1,\theta) \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1} dx$$

- Unobserved Poisson process

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\, dx \Rightarrow (1-x) \cdot \gamma\theta x^{-1}(1-x)^{\theta-1} dx$$

Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j|a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j|a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx \Rightarrow \mathrm{Beta}(x|1,\theta) \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1}dx$$

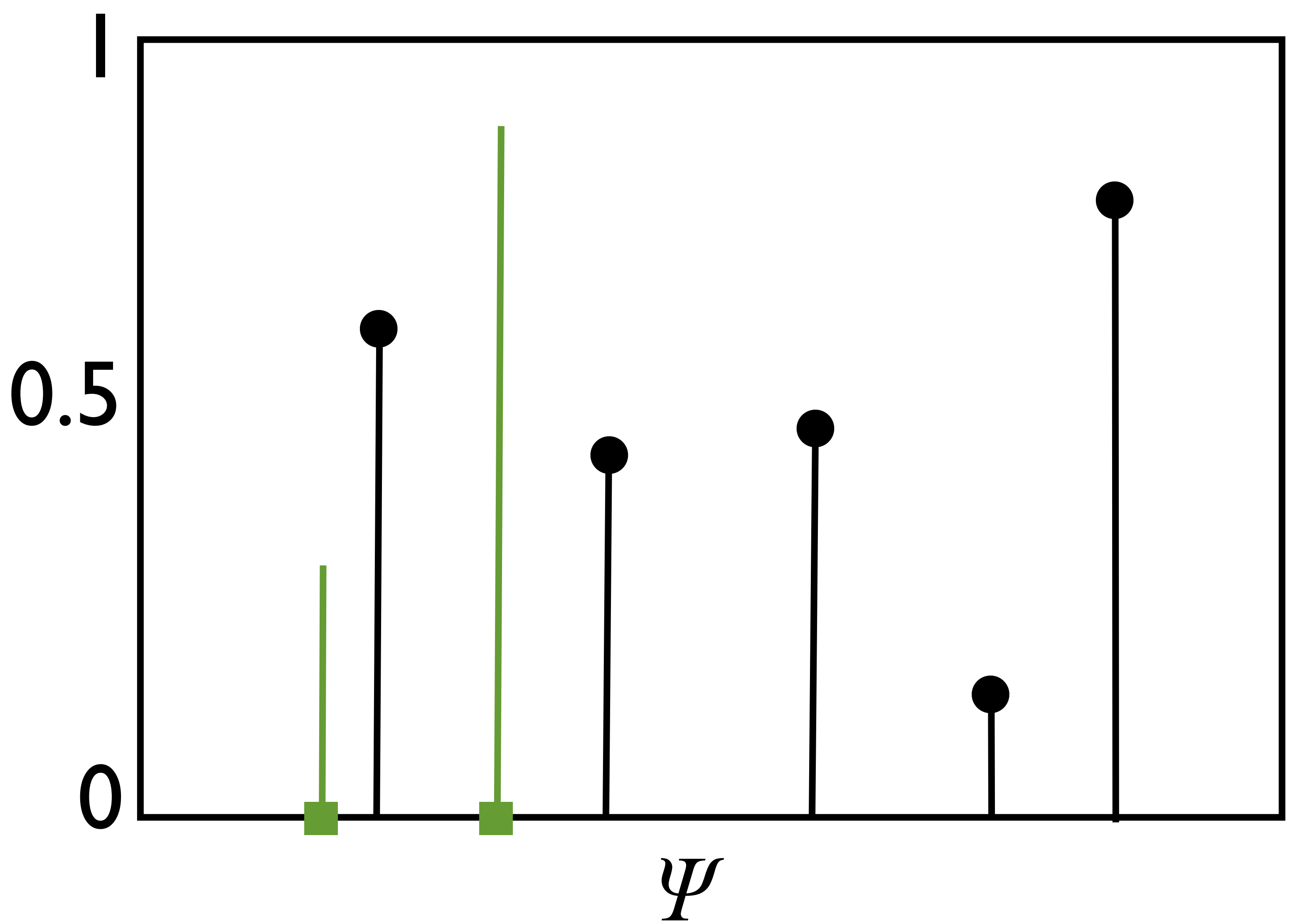
- Unobserved Poisson process

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\,dx \Rightarrow \frac{\gamma\theta}{\theta+1}(\theta+1)x^{-1}(1-x)^{(\theta+1)-1}dx$$

Completely random measure

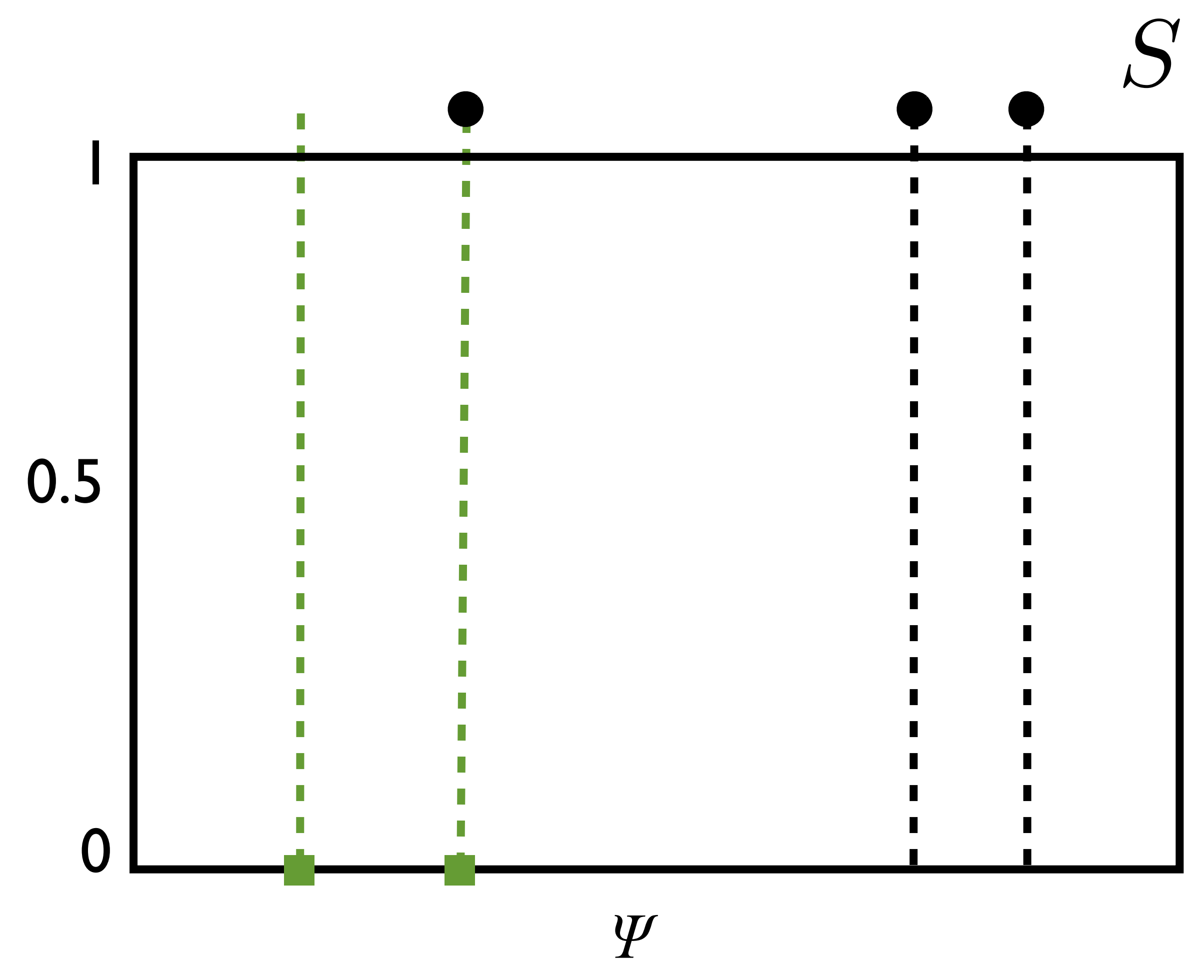
Two parts (for our purposes):

- Poisson process component
- Fixed atoms



Posterior

- Prior: beta process
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

$$\mathrm{Beta}(x_j | a_j, b_j) \Rightarrow \mathrm{Beta}\left(x_j | a_j + \mathbf{1}\{\psi_j \in S\}, b_j + \mathbf{1}\{\psi_j \notin S\}\right)$$

- New fixed atom at ψ

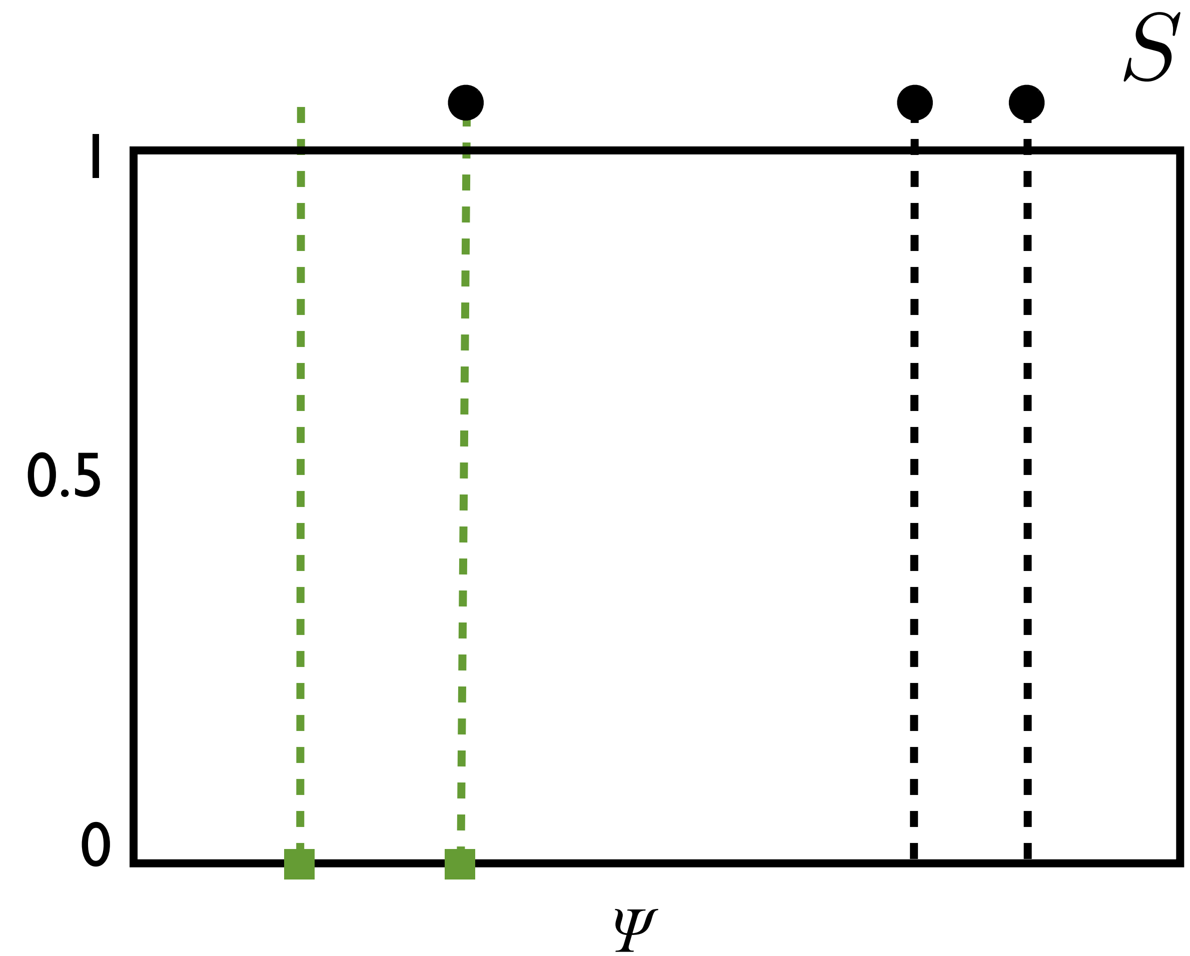
$$\gamma\theta x^{-1}(1-x)^{\theta-1}\, dx \Rightarrow \mathrm{Beta}(x|1,\theta) \propto x \cdot \gamma\theta x^{-1}(1-x)^{\theta-1}$$

- Unobserved Poisson process

$$\gamma\theta x^{-1}(1-x)^{\theta-1}\, dx \Rightarrow \frac{\gamma\theta}{\theta+1}(\theta+1)x^{-1}(1-x)^{(\theta+1)-1}$$

Posterior

- Prior: Completely random measure
- Likelihood: Bernoulli process



- Old fixed atom at ψ_j

$$\text{Usual parametric posterior}$$

- New fixed atom at ψ

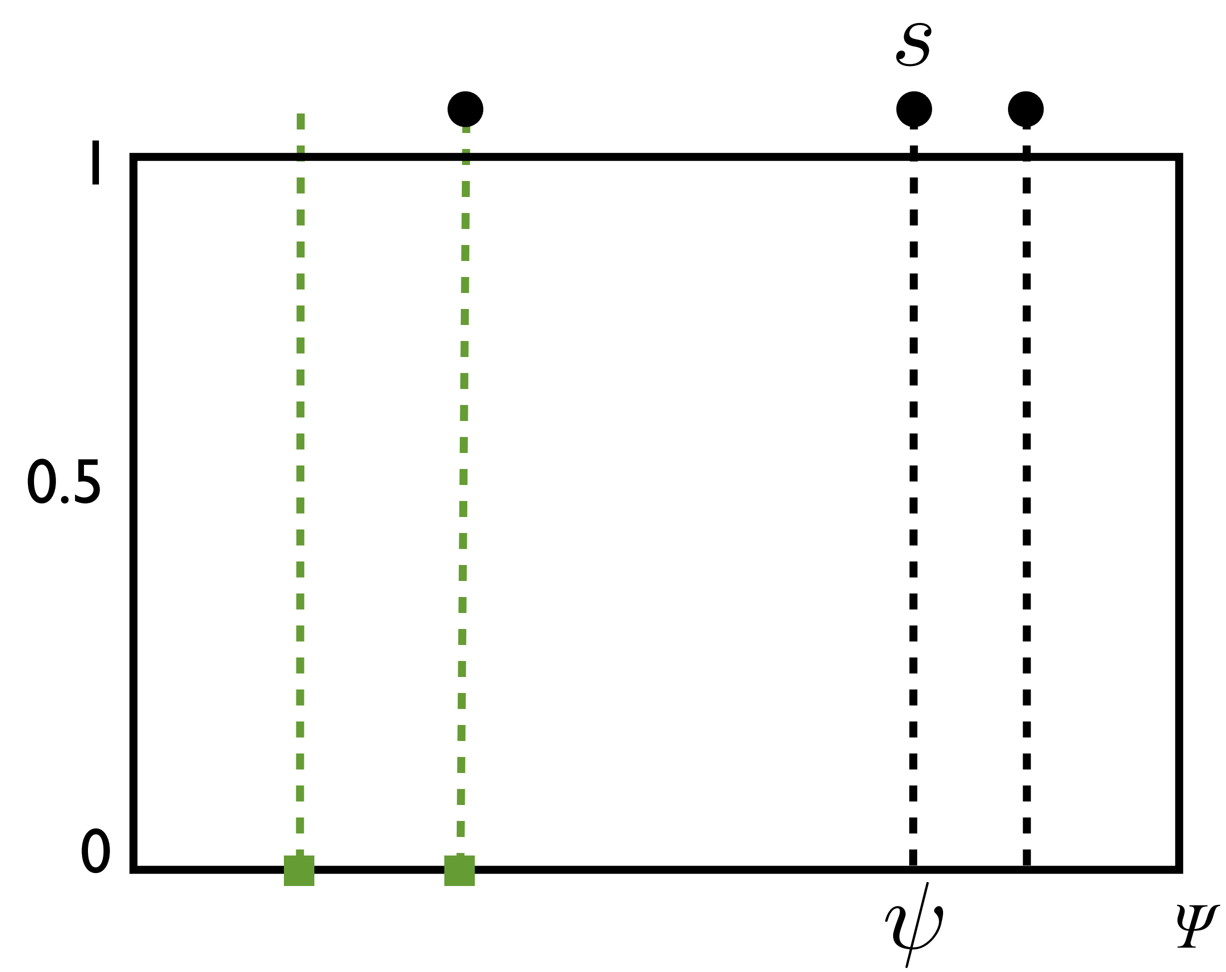
$$\nu(dx) \Rightarrow \text{density} \propto x \cdot \nu(dx)$$

- Unobserved Poisson process

$$\nu(dx) \Rightarrow (1 - x) \cdot \nu(dx)$$

Posterior

- Prior: Completely random measure
- Likelihood: Discrete process



General Recipe

- Old fixed atom at ψ_j

$$\text{Usual parametric posterior}$$

- New fixed atom at ψ

$$\nu(dx) \Rightarrow \text{density} \propto p(s|x) \cdot \nu(dx)$$

- Unobserved Poisson process

$$\nu(dx) \Rightarrow p(0|x) \cdot \nu(dx)$$

Conclusions

Conclusions

- Can identify posterior for CRM prior

Conclusions

- Can identify posterior for CRM prior
  - E.g., beta-Bernoulli

Conclusions

- Can identify posterior for CRM prior
  - E.g., beta-Bernoulli
  - And more general discrete likelihoods

Conclusions

- Can identify posterior for CRM prior
  - E.g., beta-Bernoulli
  - And more general discrete likelihoods
- Can formulate "exponential family" for CRMs

Conclusions

- Can identify posterior for CRM prior
  - E.g., beta-Bernoulli
  - And more general discrete likelihoods
- Can formulate "exponential family" for CRMs
  - And identify conjugate priors

Conclusions

- Can identify posterior for CRM prior
  - E.g., beta-Bernoulli
  - And more general discrete likelihoods
- Can formulate "exponential family" for CRMs
  - And identify conjugate priors
- Extend other ideas from parametric conjugacy

References

T. Broderick, M. I. Jordan, and J. Pitman. Clusters and features from combinatorial stochastic processes. *Statistical Science, to appear.* Preprint arXiv:1206.5862, 2012.

T. Broderick, L. Mackey, J. Paisley, and M. I. Jordan. Combinatorial clustering and the beta negative binomial process. *Preprint arXiv:1111.1802*, 2011.

T. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems*, 2006.

N. L. Hjort. Nonparametric bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, 18(3):1259–1294, 1990.

Y. Kim. Nonparametric Bayesian estimators for counting processes. *Annals of Statistics*, 27(2):562–588, 1999.

J. F. C. Kingman. *Poisson processes*. Oxford University Press, 1993.

R. Thibaux and M. I. Jordan. Hierarchical beta processes and the Indian buffet process. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2007.

M. Zhou, L. Hannah, D. Dunson, and L. Carin. Beta-negative binomial process and Poisson factor analysis. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2012.