SYDE 750 Project

Learning Probability Distributions for Statistical Inference

Andreas Stöckel

30th March, 2017

Lifespan Inference Task (I)

Based on *Optimal Predictions in Everyday Cognition* by Griffiths and Tenenbaum (2006)

- Experimental task...

  Insurance agencies employ actuaries to make predictions about people's lifespans—the age at which they will die—based upon demographic information. If you were assessing an insurance case for an 18-year-old man, what would you predict for his lifespan?

Lifespan Inference Task (I)

Based on *Optimal Predictions in Everyday Cognition* by Griffiths and Tenenbaum (2006)

- Experimental task...

  Insurance agencies employ actuaries to make predictions about people's lifespans—the age at which they will die—based upon demographic information. If you were assessing an insurance case for an 18-year-old man, what would you predict for his lifespan?

- ...as Bayesian inference

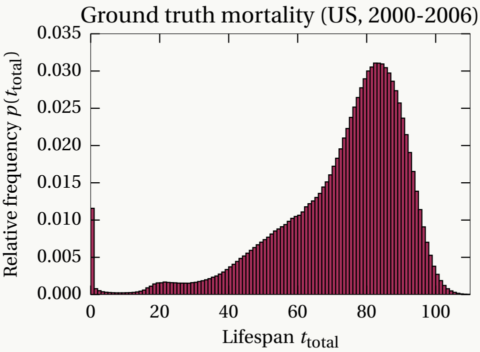
  Given t, estimate t_{total}

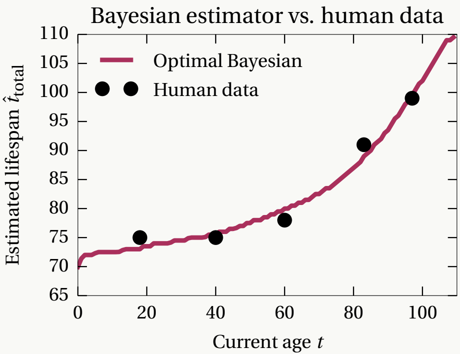
  $$p(t_{\text{total}} \mid t) = \frac{p(t \mid t_{\text{total}}) \cdot p(t_{\text{total}})}{p(t)} .$$

  Median estimator, select $\hat{t}_{\text{total}}$ s.t.

  $$p(t_{\text{total}} > \hat{t}_{\text{total}} \mid t) = p(t_{\text{total}} < \hat{t}_{\text{total}} \mid t) .$$

Lifespan Inference Task (II)



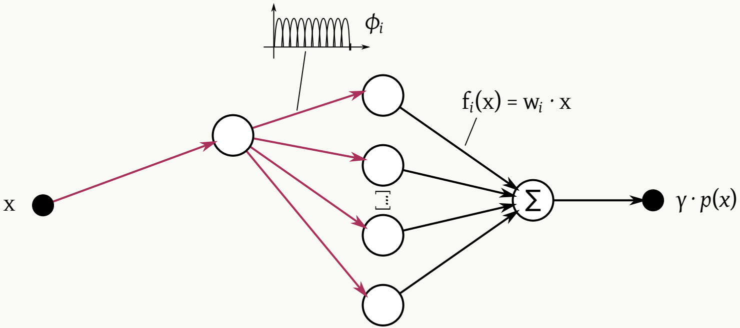


Mortality data: Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at `www.mortality.org` (data downloaded on 2017/03/29). *Human experiment results adapted from:* Optimal Predictions in Everyday Cognition. Griffiths and Tenenbaum (2006)

Learning Probability Distributions from Samples

Goal: The longer a sample x is presented, the larger $p(x)$ (without normalization).

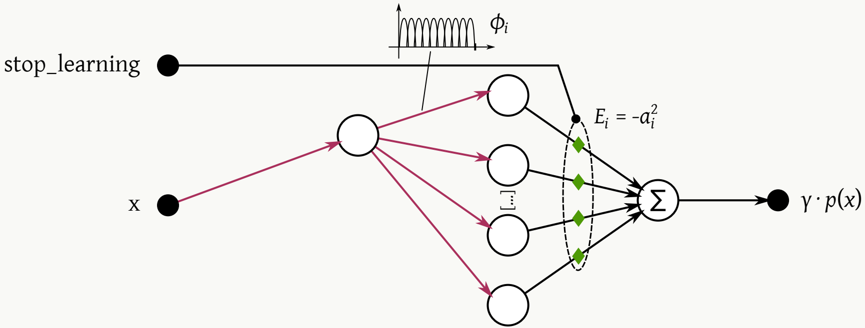
Idea: Representation as a sum of basis functions $p(x) = \sum_i w_i \cdot \phi_i(x)$.



Learning Probability Distributions from Samples

Goal: The longer a sample x is presented, the larger $p(x)$ (without normalization).

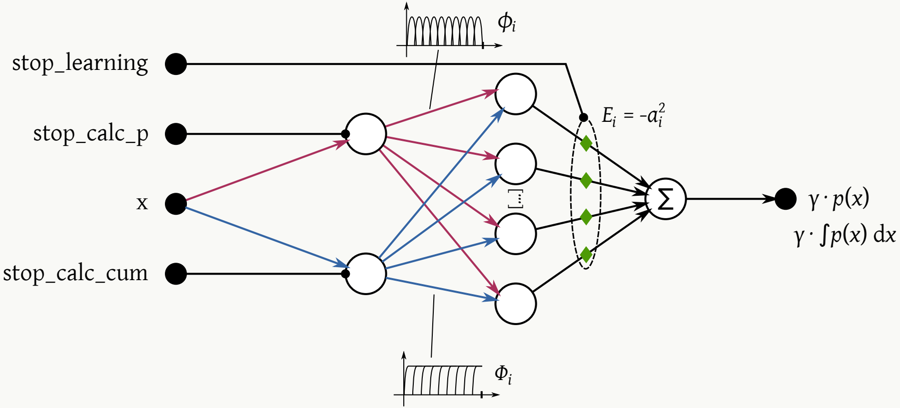
Idea: Representation as a sum of basis functions $p(x) = \sum_i w_i \cdot \phi_i(x)$.



Learning Probability Distributions from Samples

Cumulative distribution:

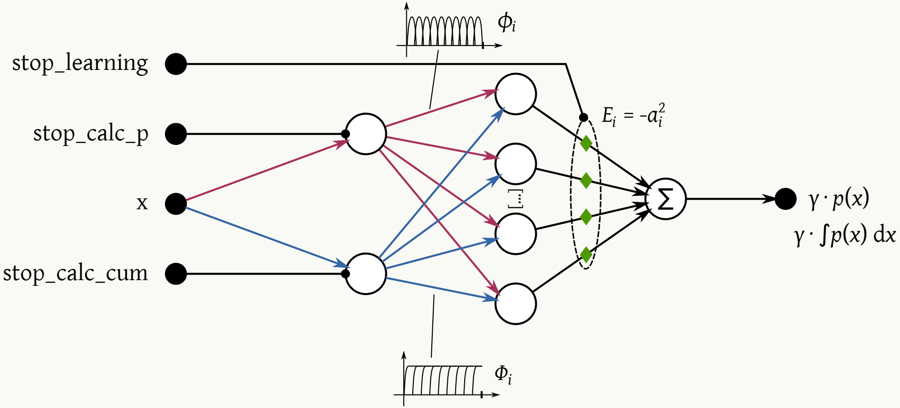
$$\int_{-\infty}^{x} p(x')\,\mathrm{d}x' = \sum_i w_i \cdot \int_{-\infty}^{x} \phi_i(x')\,\mathrm{d}x' = \sum_i w_i \cdot \Phi_i(x)$$



Learning Probability Distributions from Samples

Median estimation: Follow the gradient defined by

$$\int_{-\infty}^{x} p(x')\ \mathrm{d}x' - \int_{x}^{\infty} p(x')\ \mathrm{d}x' = \sum_i w_i \cdot \Phi_i'(x)$$



Implementing Lifespan Inference

- **Learning**

  Learn $p(t_{\text{total}})$ by sampling from the ground truth, feed samples into the network

- **Lifespan estimation**

  Calculate above median-estimating cumulative function for the posterior

  $$p(t_{\text{total}} \mid t) \propto p(t \mid t_{\text{total}}) \cdot p(t_{\text{total}}) \quad \text{where} \quad p(t \mid t_{\text{total}}) = \begin{cases} 1/t_{\text{total}} & \text{if } 0 < t < t_{\text{total}} \\ 0 & \text{otherwise} \end{cases}$$

  using a bivariate $\Phi_i(t_{\text{total}}, t)$. Integrate the output to converge to the estimate.