Restricted Boltzmann Machines: theory and applications

Gabriele Visentin [1]

March 22, 2021

[1] University of Zürich, Mathematics Institute

Focus of the talk:

- Generative models that are not so popular in particle physics.
- GANs, VAEs, etc. typically require a lot of fine tuning.
- Restricted Boltzmann Machines (RBMs):
    - are easy to train (old model $\Rightarrow$ lots of training methods!),
    - can be used for efficient conditional sampling,
    - are theoretically well studied (known density, MCMC connection).

Plan of the talk

1. Restricted Boltzmann Machines (RBMs)
    - Model description
    - Training methods
2. RBMs for credit risk management
    - Problem description
    - Model training
    - Stress testing
3. RBMs for pharmaceutical product liability
    - Problem description
    - Learning patient features
    - Learning diagnostic/clinical features
    - Legal claims distribution
    - Evaluation of alternative policies

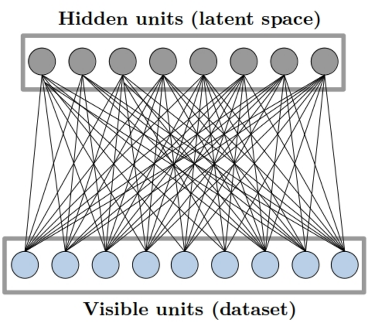
Restricted Boltzmann Machines (RBMs)

Restricted Boltzmann Machines (RBMs): Model description

An RBM is a probabilistic graphical model that can be used to learn data distributions in an unsupervised way.

Restricted Boltzmann Machines (RBMs): Model description

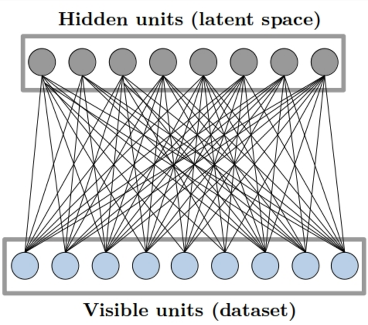
An RBM is a probabilistic graphical model that can be used to learn data distributions in an unsupervised way.



- Each vertex (or unit) corresponds to a random variable

Restricted Boltzmann Machines (RBMs): Model description

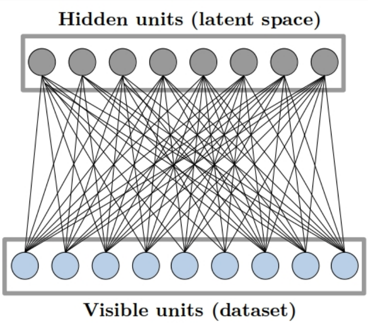
An RBM is a probabilistic graphical model that can be used to learn data distributions in an unsupervised way.



- Each vertex (or unit) corresponds to a random variable
- Units are either visible (data) or hidden (latent factors)

Restricted Boltzmann Machines (RBMs): Model description

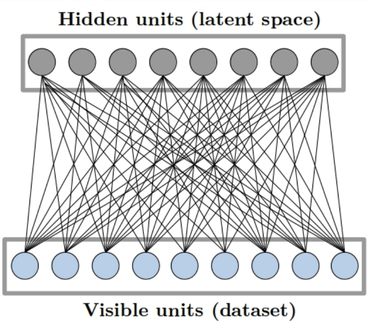
An RBM is a probabilistic graphical model that can be used to learn data distributions in an unsupervised way.



- Each vertex (or unit) corresponds to a random variable
- Units are either visible (data) or hidden (latent factors)
- Presence of an edge indicates a direct statistical dependence

Restricted Boltzmann Machines (RBMs): Model description

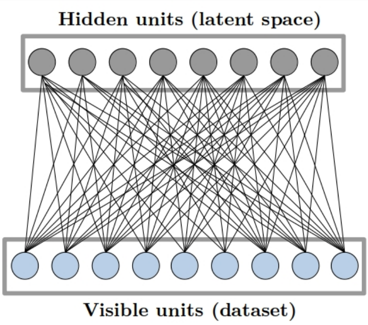
An RBM is a probabilistic graphical model that can be used to learn data distributions in an unsupervised way.



- Each vertex (or unit) corresponds to a random variable
- Units are either visible (data) or hidden (latent factors)
- Presence of an edge indicates a direct statistical dependence
- Separation relations correspond to conditional independence

Restricted Boltzmann Machines (RBMs): Model description

An RBM is a probabilistic graphical model that can be used to learn data distributions in an unsupervised way.



- Each vertex (or unit) corresponds to a random variable
- Units are either visible (data) or hidden (latent factors)
- Presence of an edge indicates a direct statistical dependence
- Separation relations correspond to conditional independence
- Hidden units are latent factors for the distribution of the visible units (non-linear version of Factor Analysis or PCA)

Restricted Boltzmann Machines (RBMs): Model description

Let's consider the case of binary units[1], i.e. $v \in \{0,1\}^n$ and $h \in \{0,1\}^m$.

[1] Generalizations to real-valued units (both visible and hidden) exist.

Restricted Boltzmann Machines (RBMs): Model description

Let's consider the case of binary units[1], i.e. $v \in \{0,1\}^n$ and $h \in \{0,1\}^m$.

An RBM is parametrized using the following Gibbs measure:

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)}$$

where

- Z is a normalization constant (partition function), such that

$$Z = \sum_{v \in \{0,1\}^n} \sum_{h \in \{0,1\}^m} e^{-E(v,h)}$$

- $E(v,h)$ is the energy function given by

$$E(v,h) = -\underbrace{\sum_{i=1}^{n} v_i b_i}_{\text{visible bias}} - \underbrace{\sum_{i=1}^{m} h_i c_i}_{\text{hidden bias}} - \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij} v_i h_j}_{\text{interaction term}}$$

[1] Generalizations to real-valued units (both visible and hidden) exist.

Restricted Boltzmann Machines (RBMs): Model description

A few preliminary remarks:

[2] Le Roux, Bengio, *Representational power of restricted Boltzmann machines and deep belief networks*, 2008

Restricted Boltzmann Machines (RBMs): Model description

A few preliminary remarks:

- The model is a universal approximator[2]. Limit case: choose as many hidden units as points in the support of the distribution. In practice: use cross-validation on the number of hidden units to avoid overfitting.

[2] Le Roux, Bengio, *Representational power of restricted Boltzmann machines and deep belief networks*, 2008

Restricted Boltzmann Machines (RBMs): Model description

A few preliminary remarks:

- The model is a universal approximator[2]. Limit case: choose as many hidden units as points in the support of the distribution. In practice: use cross-validation on the number of hidden units to avoid overfitting.
- Computing the partition function Z is numerically intractable (need to sum over 2^{n+m} terms).

[2] Le Roux, Bengio, *Representational power of restricted Boltzmann machines and deep belief networks*, 2008

Restricted Boltzmann Machines (RBMs): Model description

A few preliminary remarks:

- The model is a universal approximator[2]. Limit case: choose as many hidden units as points in the support of the distribution. In practice: use cross-validation on the number of hidden units to avoid overfitting.
- Computing the partition function Z is numerically intractable (need to sum over 2^{n+m} terms).
- Therefore computing the joint distribution $p(v, h)$ is intractable. Exact sampling from the model is not possible.

[2] Le Roux, Bengio, *Representational power of restricted Boltzmann machines and deep belief networks*, 2008

Restricted Boltzmann Machines (RBMs): Model description

A few preliminary remarks:

- The model is a universal approximator[2]. Limit case: choose as many hidden units as points in the support of the distribution. In practice: use cross-validation on the number of hidden units to avoid overfitting.
- Computing the partition function Z is numerically intractable (need to sum over 2^{n+m} terms).
- Therefore computing the joint distribution $p(v, h)$ is intractable. Exact sampling from the model is not possible.
- Nevertheless the conditional distributions $p(v|h)$ and $p(h|v)$ are easy:

$$\mathbb{P}(V_i = 1 | H = h) = \text{sigmoid}\left(\sum_{j=1}^{m} W_{ij} h_j + b_i\right)$$

$$\mathbb{P}(H_j = 1 | V = v) = \text{sigmoid}\left(\sum_{i=1}^{n} W_{ij} v_i + c_j\right)$$

[2] Le Roux, Bengio, *Representational power of restricted Boltzmann machines and deep belief networks*, 2008

Restricted Boltzmann Machines (RBMs): Training methods

We train an RBM using likelihood maximization via (stochastic) gradient ascent.

Let θ be shorthand for one of the model's parameters (W, a, b), then the log-likelihood on a sample point is:

$$\log \mathcal{L}(\theta) = \log p(v) = \log \frac{1}{Z} \sum_h e^{-E(v,h)} = \log \sum_h e^{-E(v,h)} - \log \sum_{v,h} e^{-E(v,h)}.$$

Restricted Boltzmann Machines (RBMs): Training methods

We train an RBM using likelihood maximization via (stochastic) gradient ascent.

Let θ be shorthand for one of the model's parameters (W, a, b), then the log-likelihood on a sample point is:

$$\log \mathcal{L}(\theta) = \log p(v) = \log \frac{1}{Z} \sum_h e^{-E(v,h)} = \log \sum_h e^{-E(v,h)} - \log \sum_{v,h} e^{-E(v,h)}.$$

Its derivative w.r.t θ is given by:

$$\begin{aligned} \frac{\partial}{\partial \theta} \log \mathcal{L}(\theta) &= -\sum_h p(h|v) \frac{E(v,h)}{\partial \theta} + \sum_{v,h} p(v,h) \frac{\partial E(v,h)}{\partial \theta} \\ &\approx -\mathbb{E}_{\text{data}} \left[\frac{\partial E(v,h)}{\partial \theta} \right] + \mathbb{E}_{\text{model}} \left[\frac{\partial E(v,h)}{\partial \theta} \right] \end{aligned}$$

Restricted Boltzmann Machines (RBMs): Training methods

We train an RBM using likelihood maximization via (stochastic) gradient ascent.

Let θ be shorthand for one of the model's parameters (W, a, b), then the log-likelihood on a sample point is:

$$\log \mathcal{L}(\theta) = \log p(v) = \log \frac{1}{Z} \sum_h e^{-E(v,h)} = \log \sum_h e^{-E(v,h)} - \log \sum_{v,h} e^{-E(v,h)}.$$

Its derivative w.r.t θ is given by:

$$\begin{aligned} \frac{\partial}{\partial \theta} \log \mathcal{L}(\theta) &= -\sum_h p(h|v) \frac{E(v,h)}{\partial \theta} + \sum_{v,h} p(v,h) \frac{\partial E(v,h)}{\partial \theta} \\ &\approx -\mathbb{E}_{\text{data}} \left[\frac{\partial E(v,h)}{\partial \theta} \right] + \mathbb{E}_{\text{model}} \left[\frac{\partial E(v,h)}{\partial \theta} \right] \end{aligned}$$

Sampling from the model is intractable, therefore we need an approximation of the second term.

Restricted Boltzmann Machines (RBMs): Training methods

We can sample approximately from $p(v, h)$ by performing (block) Gibbs sampling:

1. Pick $v = v_0$ from dataset.
2. Sample alternatingly $h \sim p(h|V = v)$ and $v \sim p(v|H = h)$.
3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h)$.

---

[3] Hinton, *Training products of experts by minimizing contrastive divergence*, 2002

[4] Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, 2008

[5] Salakhutdinov, *Learning in Markov random fields using tempered transitions*, 2009

Restricted Boltzmann Machines (RBMs): Training methods

We can sample approximately from $p(v, h)$ by performing (block) Gibbs sampling:

1. Pick $v = v_0$ from dataset.
2. Sample alternatingly $h \sim p(h|V = v)$ and $v \sim p(v|H = h)$.
3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h)$.

Main problems:

- Thermalization may take many sampling steps ($\geq 10^4$ for exact iid sampling).
- Equivalently, the chain may be slow-mixing.

[3] Hinton, *Training products of experts by minimizing contrastive divergence*, 2002

[4] Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, 2008

[5] Salakhutdinov, *Learning in Markov random fields using tempered transitions*, 2009

Restricted Boltzmann Machines (RBMs): Training methods

We can sample approximately from $p(v, h)$ by performing (block) Gibbs sampling:

1. Pick $v = v_0$ from dataset.
2. Sample alternatingly $h \sim p(h|V = v)$ and $v \sim p(v|H = h)$.
3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h)$.

Main problems:

- Thermalization may take many sampling steps ($\geq 10^4$ for exact iid sampling).
- Equivalently, the chain may be slow-mixing.

Main training algorithms:

- **Contrastive Divergence**[3]: even one Gibbs sampling step is enough.

[3] Hinton, *Training products of experts by minimizing contrastive divergence*, 2002

[4] Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, 2008

[5] Salakhutdinov, *Learning in Markov random fields using tempered transitions*, 2009

Restricted Boltzmann Machines (RBMs): Training methods

We can sample approximately from $p(v, h)$ by performing (block) Gibbs sampling:

1. Pick $v = v_0$ from dataset.
2. Sample alternatingly $h \sim p(h|V = v)$ and $v \sim p(v|H = h)$.
3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h)$.

Main problems:

- Thermalization may take many sampling steps ($\geq 10^4$ for exact iid sampling).
- Equivalently, the chain may be slow-mixing.

Main training algorithms:

- **Contrastive Divergence**[3]: even one Gibbs sampling step is enough.
- **Stochastic Maximum Likelihood**[4]: initialize new chain with last Gibbs state of previous chain, instead of a new training sample point.

[3] Hinton, *Training products of experts by minimizing contrastive divergence*, 2002

[4] Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, 2008

[5] Salakhutdinov, *Learning in Markov random fields using tempered transitions*, 2009

Restricted Boltzmann Machines (RBMs): Training methods

We can sample approximately from $p(v, h)$ by performing (block) Gibbs sampling:

1. Pick $v = v_0$ from dataset.
2. Sample alternatingly $h \sim p(h|V = v)$ and $v \sim p(v|H = h)$.
3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h)$.

Main problems:

- Thermalization may take many sampling steps ($\geq 10^4$ for exact iid sampling).
- Equivalently, the chain may be slow-mixing.

Main training algorithms:

- **Contrastive Divergence**[3]: even one Gibbs sampling step is enough.
- **Stochastic Maximum Likelihood**[4]: initialize new chain with last Gibbs state of previous chain, instead of a new training sample point.
- **Parallel Tempering**[5]: increase mixing rate by annealed sampling.

---

[3] Hinton, *Training products of experts by minimizing contrastive divergence*, 2002

[4] Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, 2008

[5] Salakhutdinov, *Learning in Markov random fields using tempered transitions*, 2009

Restricted Boltzmann Machines (RBMs): Training methods

We can sample approximately from $p(v, h)$ by performing (block) Gibbs sampling:

1. Pick $v = v_0$ from dataset.
2. Sample alternatingly $h \sim p(h|V = v)$ and $v \sim p(v|H = h)$.
3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h)$.

Main problems:

- Thermalization may take many sampling steps ($\geq 10^4$ for exact iid sampling).
- Equivalently, the chain may be slow-mixing.

Main training algorithms:

- **Contrastive Divergence**[3]: even one Gibbs sampling step is enough.
- **Stochastic Maximum Likelihood**[4]: initialize new chain with last Gibbs state of previous chain, instead of a new training sample point.
- **Parallel Tempering**[5]: increase mixing rate by annealed sampling.
- Other methods: Pseudo-likelihood, ratio-matching, denoising score-matching.

[3] Hinton, *Training products of experts by minimizing contrastive divergence*, 2002

[4] Tieleman, *Training restricted Boltzmann machines using approximations to the likelihood gradient*, 2008

[5] Salakhutdinov, *Learning in Markov random fields using tempered transitions*, 2009

Restricted Boltzmann Machines (RBMs): Training methods

Advantages of the model:

- Fast and easy training (e.g. compared to adversarial learning).
- Conditional sampling is a built-in feature!

  To sample (v, h) given $v_i = x$:

  1. Pick $v = v_0$ from dataset (or random).
  2. Sample $h \sim p(h|V = v)$, sample $v \sim p(v|H = h)$ and fix $v_i = x$.
  3. Repeat until Markov Chain thermalizes and you obtain $(v, h) \sim p(v, h|v_i = x)$.

RBMs for credit risk management

Joint work with Giuseppe Genovese[6] and Ashkan Nikeghbali[7] [8].

[6]University of Basel, Department of Mathematics and Computer Science

[7]UZH, Mathematics Institute

[8]UZH, Department of Banking and Finance

RBMs for credit risk: Problem description

Background information:

- We focus on credit portfolios of large firms (portfolio of loans or bonds).

RBMs for credit risk: Problem description

Background information:

- We focus on credit portfolios of large firms (portfolio of loans or bonds).
- Main problem is how to model faithfully the dependence between default probabilities.

RBMs for credit risk: Problem description

Background information:

- We focus on credit portfolios of large firms (portfolio of loans or bonds).
- Main problem is how to model faithfully the dependence between default probabilities.
- Standard credit risk models use parametric or semi-parametric models (e.g. copulas, mixture models, regression).

RBMs for credit risk: Problem description

Background information:

- We focus on credit portfolios of large firms (portfolio of loans or bonds).
- Main problem is how to model faithfully the dependence between default probabilities.
- Standard credit risk models use parametric or semi-parametric models (e.g. copulas, mixture models, regression).
- Macroeconomic variables may or may not be used as underlying factors.

RBMs for credit risk: Problem description

Background information:

- We focus on credit portfolios of large firms (portfolio of loans or bonds).
- Main problem is how to model faithfully the dependence between default probabilities.
- Standard credit risk models use parametric or semi-parametric models (e.g. copulas, mixture models, regression).
- Macroeconomic variables may or may not be used as underlying factors.

Goal:

- Model the joint distribution of default probabilities and macroeconomic factors using RBMs.

RBMs for credit risk: Problem description

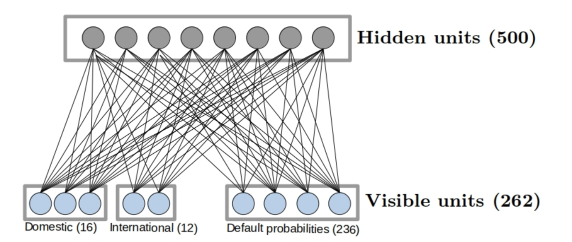
Background information:

- We focus on credit portfolios of large firms (portfolio of loans or bonds).
- Main problem is how to model faithfully the dependence between default probabilities.
- Standard credit risk models use parametric or semi-parametric models (e.g. copulas, mixture models, regression).
- Macroeconomic variables may or may not be used as underlying factors.

Goal:

- Model the joint distribution of default probabilities and macroeconomic factors using RBMs.
- Perform portfolio stress-testing (e.g. how do portfolio losses change if unemployment increases?).

RBMs for credit risk: Model training

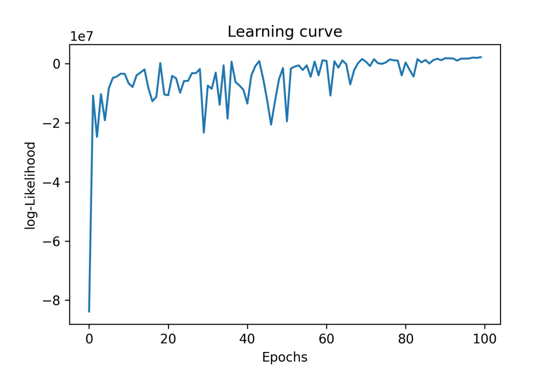


- Data.
  - Daily estimated 1-year default probabilities[9] from January 2000 to March 2020 of 236 top listed US firms.
  - Quarterly macroeconomic variables[10] (domestic and international).
- Training: hidden units: 500 (5-fold cross-validated), epochs: 10000, method: Stochastic Maximum Likelihood (100 Gibbs steps).

[9] Estimation via vanilla Merton model, similarly to Bloomberg's DRISK™ and Moody's EDF™.

[10] For a complete list see the Federal Reserve 2020Q4 stress testing documentation.

RBMs for credit risk: Model training



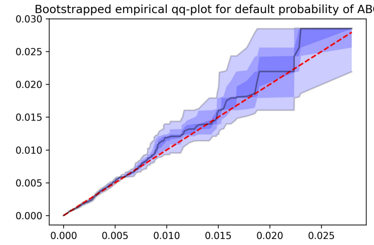
The log-likelihood is intractable (especially at training time!).

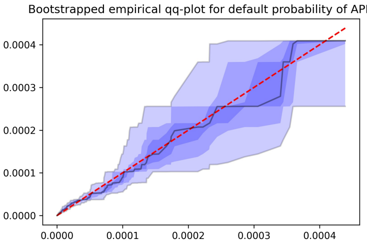
Fast ways to monitor learning:

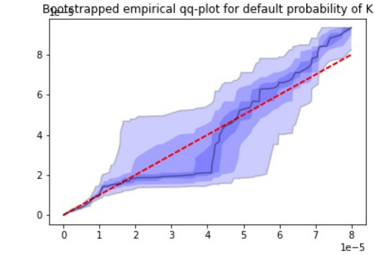
- Log-likelihood estimation via KDE from a model's sample.
- Annealed Importance Sampling for approximation of partition function[11].

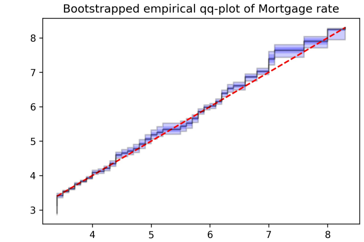
[11] See Salakhutdinov, Murray, *On the quantitative analysis of Deep Belief Networks*, 2008

RBMs for credit risk: Model training



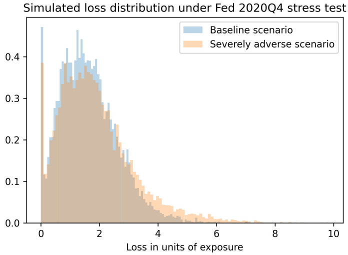


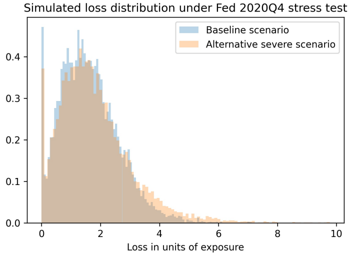




The model has successfully learned the **joint** probability distribution.

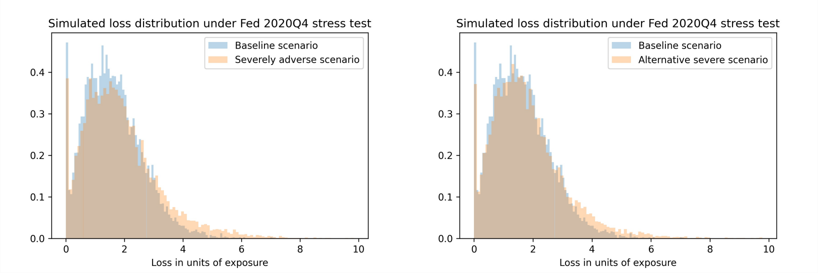
RBMs for credit risk: Stress testing





- We can implement Federal Reserve 2020Q4 stress-test by sampling conditionally on their projected macroeconomic variables and see how they affect the total losses distribution.

RBMs for credit risk: Stress testing



- We can implement Federal Reserve 2020Q4 stress-test by sampling conditionally on their projected macroeconomic variables and see how they affect the total losses distribution.
- We can compute risk measures (capital requirements) under different scenarios:
    - Value at Risk (95%)
      baseline (3.38), alternative severe (4.06, ↑ 20.1%), severely adverse (4.33, ↑ 28.1%)
    - Expected Shortfall (95%)
      baseline (4.02), alternative severe (5.15, ↑ 28.1%), severely adverse (5.37, ↑ 33.6%)

RBMs for pharmaceutical product liability

Joint work with Nicola Serra[12], Giuseppe Genovese[13], and Ashkan Nikeghbali[14] [15].

[12]UZH, Physics Institute

[13]University of Basel, Department of Mathematics and Computer Science

[14]UZH, Mathematics Institute

[15]UZH, Department of Banking and Finance

RBMs for pharma: Problem description

Background information:

- Trastuzumab is a very effective medicine used to treat a specific kind of very aggressive breast cancer (HER2-positive[16]).

[16] HER2-positive cancer is receptive to the human epidermal growth factor. Trastuzumab downregulates it, thus reducing cancer growth.

RBMs for pharma: Problem description

Background information:

- Trastuzumab is a very effective medicine used to treat a specific kind of very aggressive breast cancer (HER2-positive[16]).
- Treatment with Trastuzumab can cause cardiotoxicity (cardiac fatigue, heart failure): it is important to diagnose HER2-positivity correctly.

[16] HER2-positive cancer is receptive to the human epidermal growth factor. Trastuzumab downregulates it, thus reducing cancer growth.

RBMs for pharma: Problem description

Background information:

- Trastuzumab is a very effective medicine used to treat a specific kind of very aggressive breast cancer (HER2-positive[16]).
- Treatment with Trastuzumab can cause cardiotoxicity (cardiac fatigue, heart failure): it is important to diagnose HER2-positivity correctly.
- HER2-positivity tests: FISH (very low false positives, high cost), IHC (higher false positive rate, lower cost). Standard diagnostic strategy: IHC first, FISH only if IHC (moderately) positive.

[16] HER2-positive cancer is receptive to the human epidermal growth factor. Trastuzumab downregulates it, thus reducing cancer growth.

RBMs for pharma: Problem description

Background information:

- Trastuzumab is a very effective medicine used to treat a specific kind of very aggressive breast cancer (HER2-positive[16]).
- Treatment with Trastuzumab can cause cardiotoxicity (cardiac fatigue, heart failure): it is important to diagnose HER2-positivity correctly.
- HER2-positivity tests: FISH (very low false positives, high cost), IHC (higher false positive rate, lower cost). Standard diagnostic strategy: IHC first, FISH only if IHC (moderately) positive.

Goal:

- Learn joint distribution of patient features (e.g. age, tumor status, survival) and clinical/diagnostic features (HER2+, tests, cardiotoxicity).

[16] HER2-positive cancer is receptive to the human epidermal growth factor. Trastuzumab downregulates it, thus reducing cancer growth.

RBMs for pharma: Problem description

Background information:

- Trastuzumab is a very effective medicine used to treat a specific kind of very aggressive breast cancer (HER2-positive[16]).
- Treatment with Trastuzumab can cause cardiotoxicity (cardiac fatigue, heart failure): it is important to diagnose HER2-positivity correctly.
- HER2-positivity tests: FISH (very low false positives, high cost), IHC (higher false positive rate, lower cost). Standard diagnostic strategy: IHC first, FISH only if IHC (moderately) positive.

Goal:

- Learn joint distribution of patient features (e.g. age, tumor status, survival) and clinical/diagnostic features (HER2+, tests, cardiotoxicity).
- Model financial impact of product liability claims (legal claims due to lack of therapeutic success, serious side effects, diagnostic failure).

[16] HER2-positive cancer is receptive to the human epidermal growth factor. Trastuzumab downregulates it, thus reducing cancer growth.

RBMs for pharma: Problem description

Background information:

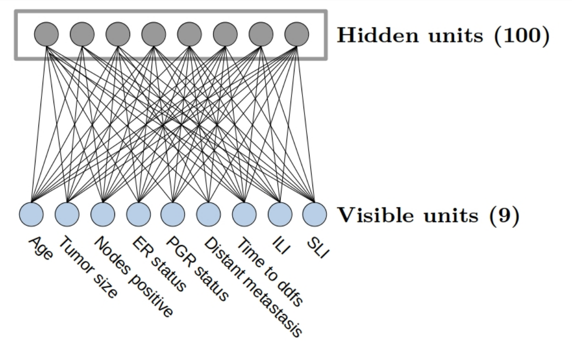
- Trastuzumab is a very effective medicine used to treat a specific kind of very aggressive breast cancer (HER2-positive[16]).
- Treatment with Trastuzumab can cause cardiotoxicity (cardiac fatigue, heart failure): it is important to diagnose HER2-positivity correctly.
- HER2-positivity tests: FISH (very low false positives, high cost), IHC (higher false positive rate, lower cost). Standard diagnostic strategy: IHC first, FISH only if IHC (moderately) positive.

Goal:

- Learn joint distribution of patient features (e.g. age, tumor status, survival) and clinical/diagnostic features (HER2+, tests, cardiotoxicity).
- Model financial impact of product liability claims (legal claims due to lack of therapeutic success, serious side effects, diagnostic failure).
- Test alternative treatments and diagnostic procedures.

[16] HER2-positive cancer is receptive to the human epidermal growth factor. Trastuzumab downregulates it, thus reducing cancer growth.

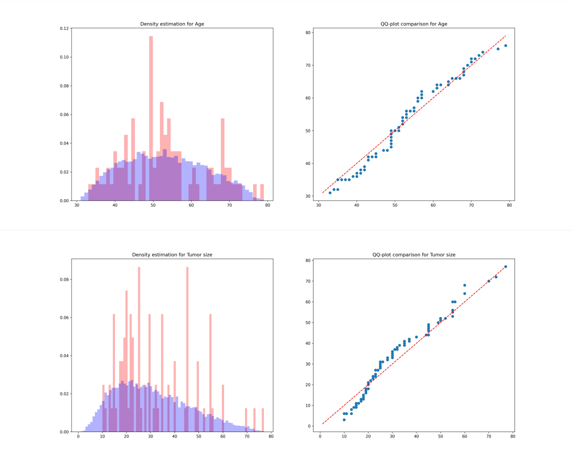
RBMs for pharma: Learning patient features



- Data: GEO2R dataset[17] with patient features from 94 HER2+ breast cancer patients[18].
- Training: hidden units: 100 (5-fold cross-validated), epochs: 10000, method: Stochastic Maximum Likelihood (100 Gibbs steps).

[17] The National Center for Biotechnology Information (NCBI) provides public access to Gene Expression Omnibus (GEO) dataset. This dataset contains gene profiling of HER2+ breast cancer patients treated with Trastuzumab.
[18] Larger datasets require long authorization procedures, in the following we will use the RBM to generate a bigger sythentic dataset on which to test our methodology.

RBMs for pharma: Learning patient features



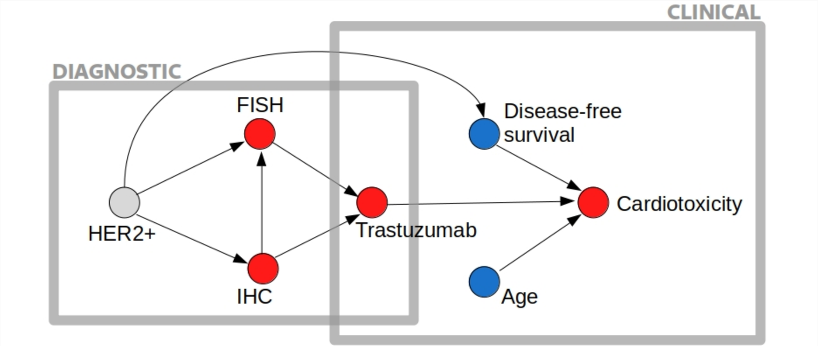
- Due to small sample size, the RBM smoothens the empirical distribution to avoid overfitting and generalize well.

RBMs for pharma: Learning diagnostic/clinical features

- We are also interested in diagnostic and clinical features not present in the GEO2R dataset.

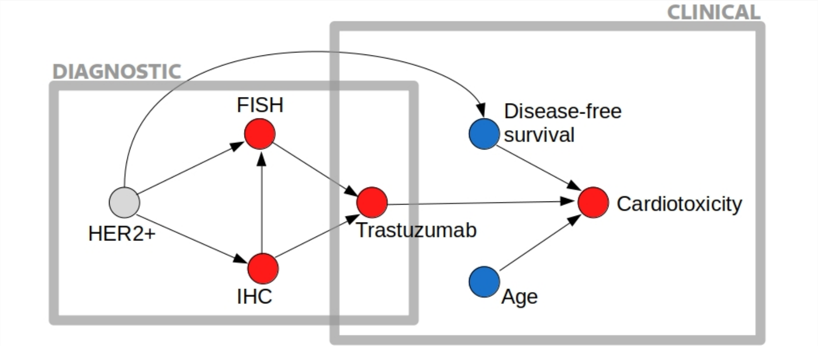
RBMs for pharma: Learning diagnostic/clinical features

- We are also interested in diagnostic and clinical features not present in the GEO2R dataset.
- Feature extension is performed using a Bayesian Network and domain-knowledge for distribution modelling (every edge is modelled using peer-reviewed research).



RBMs for pharma: Learning diagnostic/clinical features

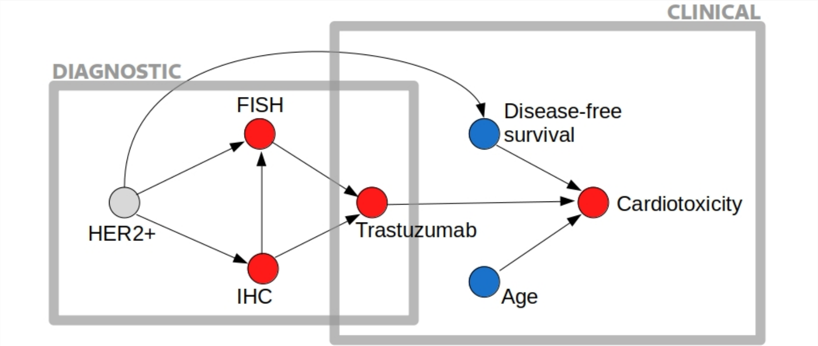
- We are also interested in diagnostic and clinical features not present in the GEO2R dataset.
- Feature extension is performed using a Bayesian Network and domain-knowledge for distribution modelling (every edge is modelled using peer-reviewed research).



- Given the value of observed nodes, we can infer the value of unobserved and unobservable nodes by sampling from the Bayesian network.

RBMs for pharma: Learning diagnostic/clinical features

- We are also interested in diagnostic and clinical features not present in the GEO2R dataset.
- Feature extension is performed using a Bayesian Network and domain-knowledge for distribution modelling (every edge is modelled using peer-reviewed research).



- Given the value of observed nodes, we can infer the value of unobserved and unobservable nodes by sampling from the Bayesian network.
- We obtain a synthetically generated sample (n=10000) from our RBM and extend it to include diagnostic/clinical variables using the Bayesian Network.

RBMs for pharma: Learning diagnostic/clinical features

We can use our model to answer basic queries:

- Frequency of Type I error (false positive) of current diagnostic strategy: 6.31%.
- Primary cardiotoxicity[19] is approx. 4.5 times more likely in 80-year-olds than 40-year-olds.
- IHC is 53% more likely than FISH to result in false positives.

What's the financial impact due to legal claims?

[19] Congestive Heart Failure or any cardiac event which may lead to hospitalization.

RBMs for pharma: Legal claims distribution

The connection between diagnostic/clinical variables and the size of legal claims might be given, for example, by the following educated guess:

Metastasis	Cardiotoxicity	Multiplier	Claim probability
Yes	Primary	1.50	50%
Yes	Secondary	1.25	35%
Yes	None	1.00	25%
No	Primary	1.00	15%
No	Secondary	0.20	5%
No	None	0.00	0%

- The expected claim size is:

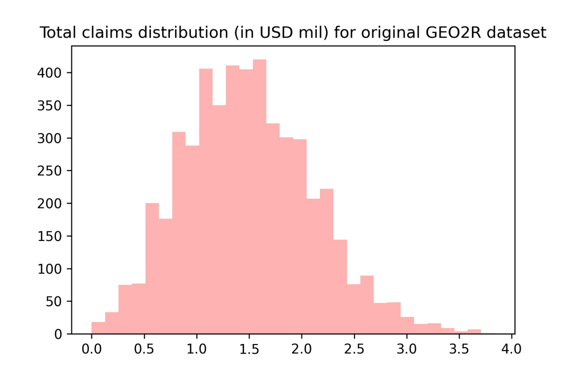
  [Median claim size = 250'000 USD] x [Multiplier] x [Claim probability]

RBMs for pharma: Legal claims distribution

- **At the beginning of treatment** given a specific group of patients (i.e. age, tumor size, nodes positive, ER status, PGR status), we can estimate the distribution of financial **losses from future claims**.

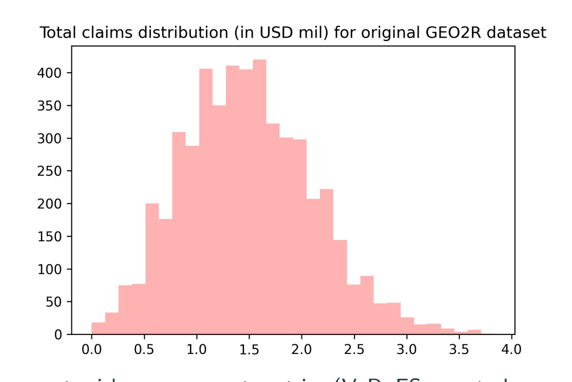
RBMs for pharma: Legal claims distribution

- **At the beginning of treatment** given a specific group of patients (i.e. age, tumor size, nodes positive, ER status, PGR status), we can estimate the distribution of financial **losses from future claims**.
- For the patients (n=94) in the original GEO2R dataset the total claims distribution is:



RBMs for pharma: Legal claims distribution

- **At the beginning of treatment** given a specific group of patients (i.e. age, tumor size, nodes positive, ER status, PGR status), we can estimate the distribution of financial **losses from future claims**.
- For the patients (n=94) in the original GEO2R dataset the total claims distribution is:



- We can compute risk management metrics (VaR, ES, spectral measures).
- VaR(95%) = 2.56 USD mil, VaR(99%) = 3.10 USD mil.
- ES(95%) = 2.88 USD mil, ES(99%) = 3.60 USD mil.

RBMs for pharma: Evaluation of alternative policies

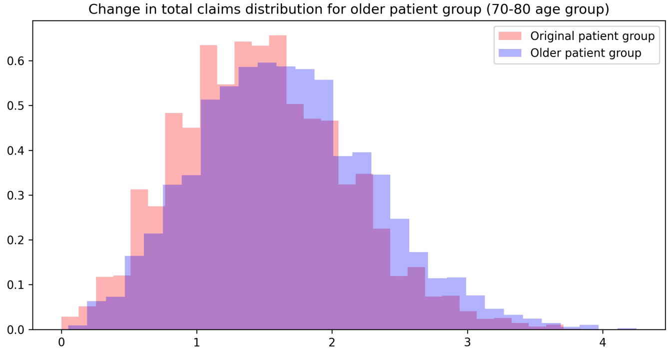
- We can estimate conditional distributions and evaluate **how clinical changes affect the distribution of losses**.

RBMs for pharma: Evaluation of alternative policies

- We can estimate conditional distributions and evaluate **how clinical changes affect the distribution of losses**.
- What is the **effect of patient age on the total claims** distribution? We sample from the model, given that all patients are in the 70-80 age group.

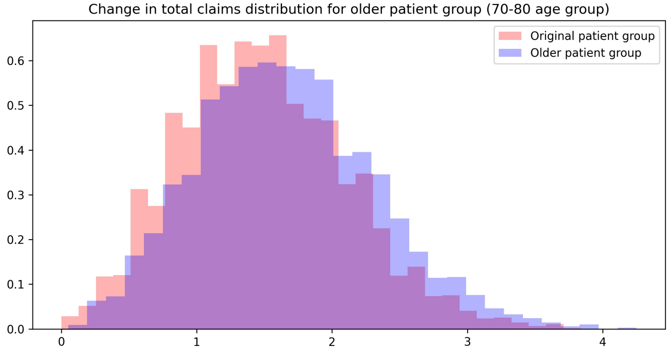
RBMs for pharma: Evaluation of alternative policies

- We can estimate conditional distributions and evaluate **how clinical changes affect the distribution of losses**.
- What is the **effect of patient age on the total claims** distribution? We sample from the model, given that all patients are in the 70-80 age group.



RBMs for pharma: Evaluation of alternative policies

- We can estimate conditional distributions and evaluate **how clinical changes affect the distribution of losses**.
- What is the **effect of patient age on the total claims** distribution? We sample from the model, given that all patients are in the 70-80 age group.



- VaR(95%) = 2.80 USD mil (↑ 9.27%)
- ES(95%) = 3.13 USD mil (↑ 8.47%)

RBMs for pharma: Evaluation of alternative policies

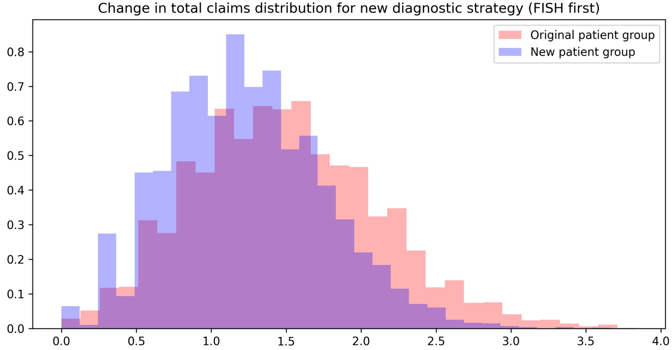
- We can evaluate the **effect of alternative diagnostic strategies on the total claims** distribution.

RBMs for pharma: Evaluation of alternative policies

- We can evaluate the **effect of alternative diagnostic strategies on the total claims** distribution.
- We sample from the model, assuming an alternative diagnostic procedure (for example: always use FISH test first).

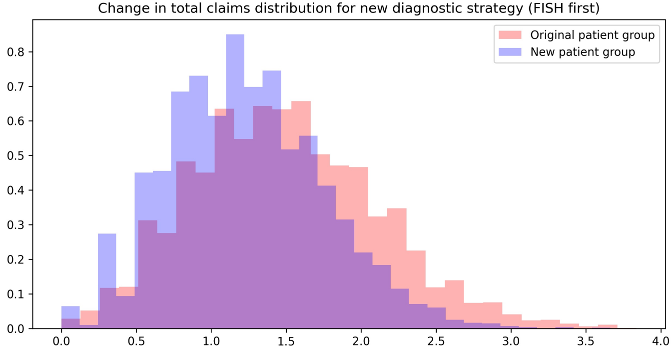
RBMs for pharma: Evaluation of alternative policies

- We can evaluate the **effect of alternative diagnostic strategies on the total claims** distribution.
- We sample from the model, assuming an alternative diagnostic procedure (for example: always use FISH test first).



RBMs for pharma: Evaluation of alternative policies

- We can evaluate the **effect of alternative diagnostic strategies on the total claims** distribution.
- We sample from the model, assuming an alternative diagnostic procedure (for example: always use FISH test first).



- VaR(95%) = 2.13 USD mil (↓ 17.1 %)
- ES(95%) = 2.39 USD mil (↓ 17.2 %)

RBMs for pharma: Evaluation of alternative policies

This methodology allows us to:

- Augment and extend existing datasets.
- Combine peer-reviewed research and ML.
- Generate synthetic datasets exhibiting complex non-linear dependencies.
- Estimate quantitatively how different diagnostic/clinical strategies can impact financial losses due to liability claims.

Conclusions

- Results of the applications are preliminary: papers are still in progress.
- RBMs are easy to train and easy to deploy.
- Conditional sampling is efficient and very useful for scenario generation.

Thanks for your attention!