# Learning the latent structure of background distributions

**Santiago Tanco**

INFN Genova

In collaboration with Ezequiel Álvarez[1] and Jernej Kamenik[2]

[1]ICAS, CONICET, Argentina

[2]Jožef Stefan Institute, Slovenia

ML4Jets - Vienna

September 17, 2026

# Challenges in background estimation

- Accurate background estimation is crucial to any experimental analysis

# Challenges in background estimation

- Accurate background estimation is crucial to any experimental analysis
- **Monte Carlo simulations**
  - Encode our knowledge of known physics
  - May not accurately reproduce data in some situations
  - Depend on a specific parametrization

# Challenges in background estimation

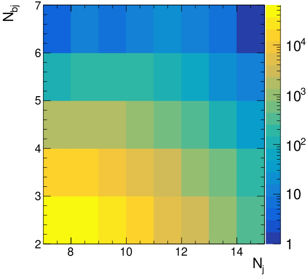
- Accurate background estimation is crucial to any experimental analysis
- **Monte Carlo simulations**
  - Encode our knowledge of known physics
  - May not accurately reproduce data in some situations
  - Depend on a specific parametrization
- **Data-driven methods**
  - ABCD and similar techniques extrapolate from background-dominated regions
  - Key assumptions*: observable independence, signal confinement
  - *may not be exactly fulfilled, resulting in biased estimations

# Challenges in background estimation

- Accurate background estimation is crucial to any experimental analysis
- **Monte Carlo simulations**
  - Encode our knowledge of known physics
  - May not accurately reproduce data in some situations
  - Depend on a specific parametrization
- **Data-driven methods**
  - ABCD and similar techniques extrapolate from background-dominated regions
  - Key assumptions*: observable independence, signal confinement
  - *may not be exactly fulfilled, resulting in biased estimations

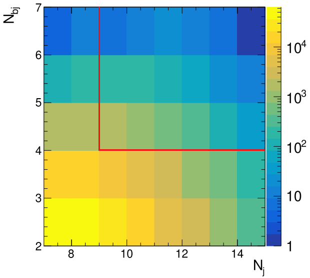
**How can we inject more information from known physics (multiple MC samples) into a data-driven inference pipeline?**

# A practical example: $pp \to t\bar{t}jj$




S. Choi and H. Oh [1906.10831]

# A practical example: $pp \to t\bar{t}jj$



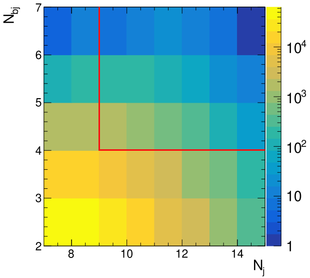

S. Choi and H. Oh [1906.10831]

Data-driven background estimation:

- The expected signal lies in a specific region (SR) in the space of 2 observables
- The **correlation** between observables is small

$$P(x,y) = P_x(x)\, P_y(y)\, \left[1 + \epsilon(x,y)\right]$$

# A practical example: $pp \to t\bar{t}jj$




S. Choi and H. Oh [1906.10831]

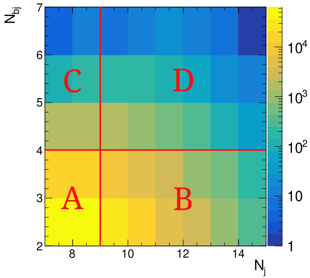
Data-driven background estimation:

- The expected signal lies in a specific region (SR) in the space of 2 observables
- The **correlation** between observables is small

$$P(x,y) = P_x(x)\, P_y(y)\; [1 + \epsilon(x,y)]$$

- Use event counts in control regions (CR) to estimate the background in SR

# A practical example: $pp \to t\bar{t}jj$




S. Choi and H. Oh [1906.10831]

- If $\epsilon = 0$,

$$\hat{D} = \frac{C}{A} \times B$$

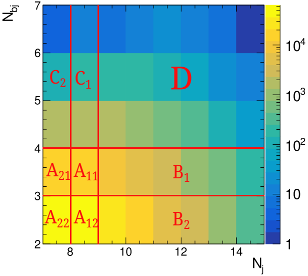
Data-driven background estimation:

- The expected signal lies in a specific region (SR) in the space of 2 observables
- The **correlation** between observables is small

$$P(x,y) = P_x(x)\, P_y(y)\ [1 + \epsilon(x,y)]$$

- Use event counts in control regions (CR) to estimate the background in SR

# A practical example: $pp \to t\bar{t}jj$




S. Choi and H. Oh [1906.10831]

- If $\epsilon = 0$,

$$\hat{D} = \frac{C}{A} \times B$$

- If $\epsilon \ll 1$,

$$\hat{D} = \left(\frac{C_2 B_2}{A_{22}}\right)^{-\frac{1}{3}} \left(\frac{C_1 B_1}{A_{11}}\right)^{\frac{4}{3}} + \mathcal{O}(\Delta^3)$$

$$\hat{D} = \frac{C_2 B_2}{A_{22}} \left(\frac{C_1 B_1}{A_{11}}\right)^4 \left(\frac{C_2 B_1}{A_{21}}\right)^{-2} \left(\frac{C_1 B_2}{A_{12}}\right)^{-2} + \mathcal{O}(\Delta^4)$$

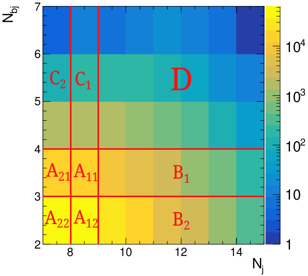
Data-driven background estimation:

- The expected signal lies in a specific region (SR) in the space of 2 observables
- The **correlation** between observables is small

$$P(x,y) = P_x(x)\, P_y(y)\; [1 + \epsilon(x,y)]$$

- Use event counts in control regions (CR) to estimate the background in SR

# A practical example: $pp \to t\bar{t}jj$




S. Choi and H. Oh [1906.10831]

- If $\epsilon = 0$,

$$\hat{D} = \frac{C}{A} \times B$$

- If $\epsilon \ll 1$,

$$\hat{D} = \left(\frac{C_2 B_2}{A_{22}}\right)^{-\frac{1}{3}} \left(\frac{C_1 B_1}{A_{11}}\right)^{\frac{4}{3}} + \mathcal{O}(\Delta^3)$$

$$\hat{D} = \frac{C_2 B_2}{A_{22}} \left(\frac{C_1 B_1}{A_{11}}\right)^4 \left(\frac{C_2 B_1}{A_{21}}\right)^{-2} \left(\frac{C_1 B_2}{A_{12}}\right)^{-2} + \mathcal{O}(\Delta^4)$$

| Extrapolation method | Prediction ($\hat{F}_D$) | $\hat{F}_D / F_D$ |
|---|---|---|
| ABCD (Eq. 12) | $3333 \pm 77$ | $0.684 \pm 0.015$ |
| Ext. ABCD (Eq. 14) | $4149 \pm 132$ | $0.851 \pm 0.027$ |
| Ext. ABCD (Eq. 15) | $4352 \pm 271$ | $0.893 \pm 0.056$ |
| Ext. ABCD (Eq. 16) | $4247 \pm 217$ | $0.871 \pm 0.045$ |

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

- **Topic**: probability distribution over words

$$\vec{\varphi}^{(k)} = (\varphi_1^{(k)}, \ldots, \varphi_V^{(k)}),$$

$$\sum_{v=1}^{V} \varphi_v^{(k)} = 1, \quad k = 1, \ldots, K$$

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

- **Topic**: probability distribution over words

$$\vec{\varphi}^{(k)} = (\varphi_1^{(k)}, \ldots, \varphi_V^{(k)}),$$

$$\sum_{v=1}^{V} \varphi_v^{(k)} = 1, \quad k = 1, \ldots, K$$

- **Document** (bag of words): mixture of topics

$$\vec{\theta}^{(d)} = (\theta_1^{(d)}, \ldots, \theta_K^{(d)}),$$

$$\sum_{k=1}^{K} \theta_k^{(d)} = 1, \quad d = 1, \ldots, M$$

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

- **Topic**: probability distribution over words

$$\vec{\varphi}^{(k)} = (\varphi_1^{(k)}, \ldots, \varphi_V^{(k)}),$$

$$\sum_{v=1}^{V} \varphi_v^{(k)} = 1, \quad k = 1, \ldots, K$$

- **Document** (bag of words): mixture of topics

$$\vec{\theta}^{(d)} = (\theta_1^{(d)}, \ldots, \theta_K^{(d)}),$$

$$\sum_{k=1}^{K} \theta_k^{(d)} = 1, \quad d = 1, \ldots, M$$

- **Generative process**: for each document $d = 1, \ldots, M$:
  - For each word $n = 1, \ldots, N_d$:
    1. Sample a topic $z_{dn} \sim \text{Categorical}(\vec{\theta}^{(d)})$
    2. Sample a word $w_{dn} \sim \text{Categorical}(\vec{\varphi}^{(z_{dn})})$

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

- **Topic**: probability distribution over words

$$\vec{\varphi}^{(k)} = (\varphi_1^{(k)}, \ldots, \varphi_V^{(k)}),$$

$$\sum_{v=1}^{V} \varphi_v^{(k)} = 1, \quad k = 1, \ldots, K$$

- **Document** (bag of words): mixture of topics

$$\vec{\theta}^{(d)} = (\theta_1^{(d)}, \ldots, \theta_K^{(d)}),$$

$$\sum_{k=1}^{K} \theta_k^{(d)} = 1, \quad d = 1, \ldots, M$$

- **Generative process**: for each document $d = 1, \ldots, M$:
  - For each word $n = 1, \ldots, N_d$:
    1. Sample a topic $z_{dn} \sim \text{Categorical}(\vec{\theta}^{(d)})$
    2. Sample a word $w_{dn} \sim \text{Categorical}(\vec{\varphi}^{(z_{dn})})$
- **Inference**: given documents $\{w_{dn}\}$, learn the topics $\vec{\varphi}^{(k)}$ and their distributions $\vec{\theta}^{(d)}$

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

- **Topic**: probability distribution over words

$$\vec{\varphi}^{(k)} = (\varphi_1^{(k)}, \ldots, \varphi_V^{(k)}),$$

$$\sum_{v=1}^{V} \varphi_v^{(k)} = 1, \quad k = 1, \ldots, K$$

- **Document** (bag of words): mixture of topics

$$\vec{\theta}^{(d)} = (\theta_1^{(d)}, \ldots, \theta_K^{(d)}),$$

$$\sum_{k=1}^{K} \theta_k^{(d)} = 1, \quad d = 1, \ldots, M$$

- **Generative process**: for each document $d = 1, \ldots, M$:
  - For each word $n = 1, \ldots, N_d$:
    1. Sample a topic $z_{dn} \sim \text{Categorical}(\vec{\theta}^{(d)})$
    2. Sample a word $w_{dn} \sim \text{Categorical}(\vec{\varphi}^{(z_{dn})})$
- **Inference**: given documents $\{w_{dn}\}$, learn the topics $\vec{\varphi}^{(k)}$ and their distributions $\vec{\theta}^{(d)}$
- **Co-ocurrence** of words tend to group them by meaning

# Topic modeling: *Latent Dirichlet Allocation* (LDA)

Given a *corpus* of $M$ **documents**, learn a representation in terms of a reduced number of $K$ topics

- **Topic**: probability distribution over words

$$\vec{\varphi}^{(k)} = (\varphi_1^{(k)}, \ldots, \varphi_V^{(k)}),$$

$$\sum_{v=1}^{V} \varphi_v^{(k)} = 1, \quad k = 1, \ldots, K$$

- **Document** (bag of words): mixture of topics

$$\vec{\theta}^{(d)} = (\theta_1^{(d)}, \ldots, \theta_K^{(d)}),$$

$$\sum_{k=1}^{K} \theta_k^{(d)} = 1, \quad d = 1, \ldots, M$$

- **Generative process**: for each document $d = 1, \ldots, M$:
  - For each word $n = 1, \ldots, N_d$:
    1. Sample a topic $z_{dn} \sim \text{Categorical}(\vec{\theta}^{(d)})$
    2. Sample a word $w_{dn} \sim \text{Categorical}(\vec{\varphi}^{(z_{dn})})$
- **Inference**: given documents $\{w_{dn}\}$, learn the topics $\vec{\varphi}^{(k)}$ and their distributions $\vec{\theta}^{(d)}$
- **Co-ocurrence** of words tend to group them by meaning $\Rightarrow$ **correlations ?**

# Using LDA in $t\bar{t}jj$

| | | |
|---|---|---|
| Corpus | $\Rightarrow$ | Set of simulations |
| Document | $\Rightarrow$ | Simulation with a given setup |
| Word | $\Rightarrow$ | Event, joint observation of $(N_j, N_b)$ |
| Topic | $\Rightarrow$ | Probability mass function over events |

## Using LDA in $t\bar{t}jj$

| | | |
|---|---|---|
| Corpus | $\Rightarrow$ | Set of simulations |
| Document | $\Rightarrow$ | Simulation with a given setup |
| Word | $\Rightarrow$ | Event, joint observation of $(N_j, N_b)$ |
| Topic | $\Rightarrow$ | Probability mass function over events |

- Learn topics as underlying correlation structures between observations
- Dimensionality reduction

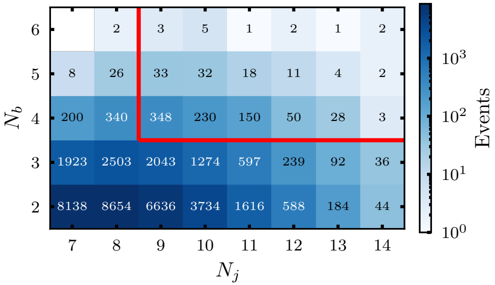
# Using LDA in $t\bar{t}jj$

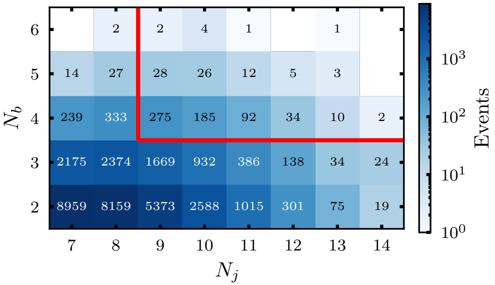
| | | |
|---|---|---|
| Corpus | $\Rightarrow$ | Set of simulations |
| Document | $\Rightarrow$ | Simulation with a given setup |
| Word | $\Rightarrow$ | Event, joint observation of $(N_j, N_b)$ |
| Topic | $\Rightarrow$ | Probability mass function over events |

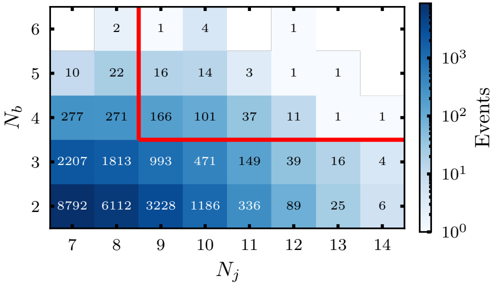
- Learn topics as underlying correlation structures between observations
- Dimensionality reduction
- Generate 70 simulations of $pp \to t\bar{t}jj$ varying factorization scale and $p_T^j$ cut

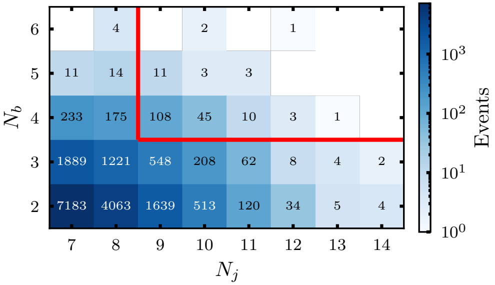
1. Learn the topics distributions from simulations
2. Once the topics are known, learn their mixing fractions in the target dataset

# Simulations of $pp \to t\bar{t}jj$



# Bayesian inference

Observations $\mathbf{W} = \{w_{dn}\}$, parameters of interest $\boldsymbol{\Theta} = \{\vec{\theta}^{(d)}\}$, $\boldsymbol{\Phi} = \{\vec{\varphi}^{(k)}\}$, probabilistic model

$$p\left(\mathbf{W} \,\middle|\, \boldsymbol{\Theta}, \boldsymbol{\Phi}\right)$$

Bayes' theorem allows us to find the **posterior** distribution,

$$\log p\left(\boldsymbol{\Theta}, \boldsymbol{\Phi} \,\middle|\, \mathbf{W}\right) = \log p\left(\mathbf{W} \,\middle|\, \boldsymbol{\Theta}, \boldsymbol{\Phi}\right) + \log p(\boldsymbol{\Theta}, \boldsymbol{\Phi}) - C$$

## Bayesian inference

Observations $\mathbf{W} = \{w_{dn}\}$, parameters of interest $\boldsymbol{\Theta} = \{\vec{\theta}^{(d)}\}$, $\boldsymbol{\Phi} = \{\vec{\varphi}^{(k)}\}$, probabilistic model

$$p\left(\mathbf{W} \,\middle|\, \boldsymbol{\Theta}, \boldsymbol{\Phi}\right)$$

Bayes' theorem allows us to find the **posterior** distribution,

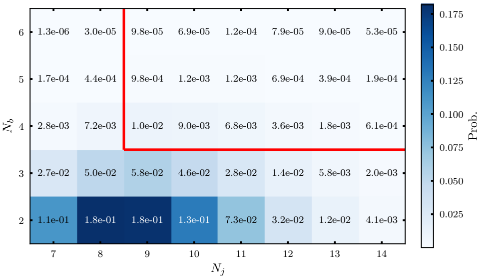
$$\log p\left(\boldsymbol{\Theta}, \boldsymbol{\Phi} \,\middle|\, \mathbf{W}\right) = \log p\left(\mathbf{W} \,\middle|\, \boldsymbol{\Theta}, \boldsymbol{\Phi}\right) + \log p(\boldsymbol{\Theta}, \boldsymbol{\Phi}) - C$$

*Hamiltonian Monte Carlo* (HMC) to sample the posterior

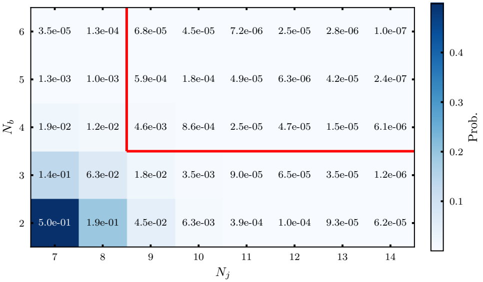
- Mapping the inference problem to a Hamiltonian dynamical system
- Parameters to infer $\longleftrightarrow$ generalized coordinates
- Auxiliary momenta $\boldsymbol{\rho}$

$$H(\boldsymbol{\rho}, \boldsymbol{\Theta}, \boldsymbol{\Phi}) = -\log p\left(\boldsymbol{\rho} \,\middle|\, \boldsymbol{\Theta}, \boldsymbol{\Phi}\right) - \log p\left(\boldsymbol{\Theta}, \boldsymbol{\Phi} \,\middle|\, \mathbf{W}\right) = T + V$$

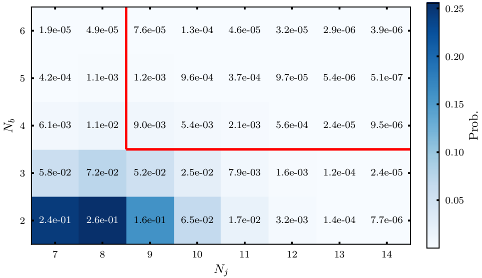
# Topics ($K = 3$)



6 1.3e-06 3.0e-05 9.8e-05 6.9e-05 1.2e-04 7.9e-05 9.0e-05 5.3e-05
5 1.7e-04 4.4e-04 9.8e-04 1.2e-03 1.2e-03 6.9e-04 3.9e-04 1.9e-04
4 2.8e-03 7.2e-03 1.0e-02 9.0e-03 6.8e-03 3.6e-03 1.8e-03 6.1e-04
3 2.7e-02 5.0e-02 5.8e-02 4.6e-02 2.8e-02 1.4e-02 5.8e-03 2.0e-03
2 1.1e-01 1.8e-01 1.8e-01 1.3e-01 7.3e-02 3.2e-02 1.2e-02 4.1e-03
7 8 9 10 11 12 13 14
$N_b$
$N_j$
0.175
0.150
0.125
0.100
0.075
0.050
0.025
Prob.

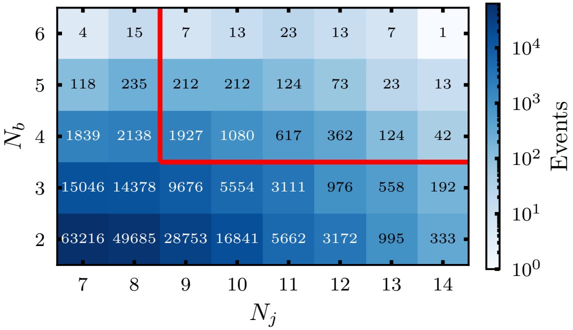



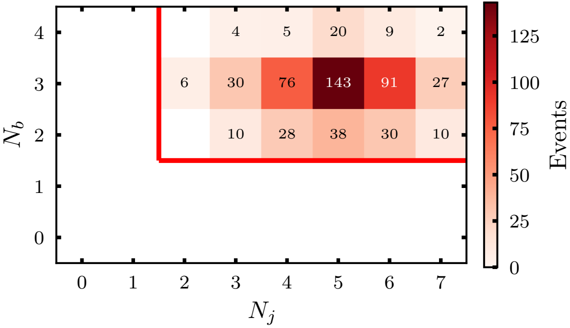
6 3.5e-05 1.3e-04 6.8e-05 4.5e-05 7.2e-06 2.5e-05 2.8e-06 1.0e-07
5 1.3e-03 1.0e-03 5.9e-04 1.8e-04 4.9e-05 6.3e-06 4.2e-05 2.4e-07
4 1.9e-02 1.2e-02 4.6e-03 8.6e-04 2.5e-05 4.7e-05 1.5e-05 6.1e-06
3 1.4e-01 6.3e-02 1.8e-02 3.5e-03 9.0e-05 6.5e-05 3.5e-05 1.2e-06
2 5.0e-01 1.9e-01 4.5e-02 6.3e-03 3.9e-04 1.0e-04 9.3e-05 6.2e-05
7 8 9 10 11 12 13 14
$N_b$
$N_j$
0.4
0.3
0.2
0.1
Prob.

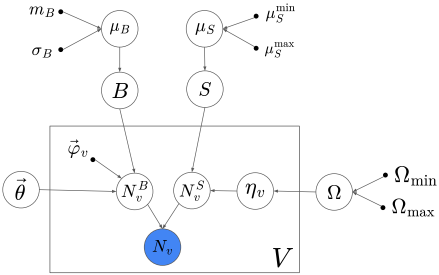
6 1.9e-05 4.9e-05 7.6e-05 1.3e-04 4.6e-05 3.2e-05 2.9e-06 3.9e-06
5 4.2e-04 1.1e-03 1.2e-03 9.6e-04 3.7e-04 9.7e-05 5.4e-06 5.1e-07
4 6.1e-03 1.1e-02 9.0e-03 5.4e-03 2.1e-03 5.6e-04 2.4e-05 9.5e-06
3 5.8e-02 7.2e-02 5.2e-02 2.5e-02 7.9e-03 1.6e-03 1.2e-04 2.4e-05
2 2.4e-01 2.6e-01 1.6e-01 6.5e-02 1.7e-02 3.2e-03 1.4e-04 7.7e-06
7 8 9 10 11 12 13 14
$N_b$
$N_j$
0.25
0.20
0.15
0.10
0.05
Prob.

# Adding signal to target data



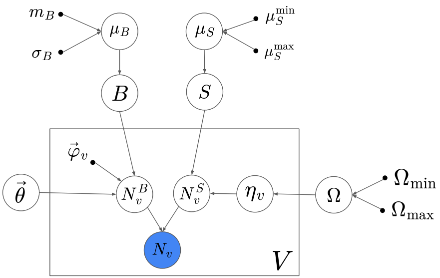
# Adding signal to the model




- Signal and background mixture,

$$S \sim \text{Poisson}(\mu_S), \quad B \sim \text{Poisson}(\mu_B)$$

# Adding signal to the model

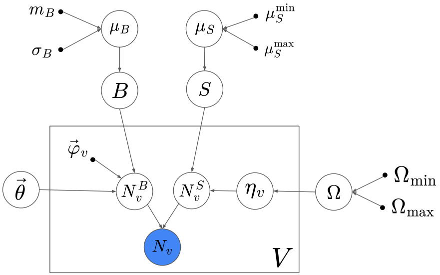



- Signal and background mixture,

$$S \sim \text{Poisson}(\mu_S), \quad B \sim \text{Poisson}(\mu_B)$$

- In CR: $N_\nu \sim \text{Poisson}\left(\mu_B \vec{\theta} \cdot \vec{\varphi}_\nu\right)$
- In SR: $N_\nu \sim \text{Poisson}\left(\mu_B \vec{\theta} \cdot \vec{\varphi}_\nu + \mu_S \eta_\nu\right)$

# Adding signal to the model

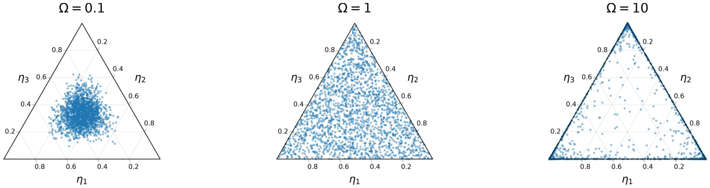



- Signal and background mixture,

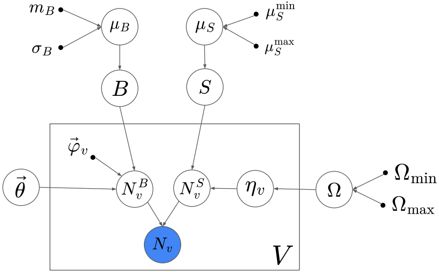
$$S \sim \text{Poisson}(\mu_S), \quad B \sim \text{Poisson}(\mu_B)$$

- In CR: $N_v \sim \text{Poisson}\left(\mu_B \vec{\theta} \cdot \vec{\varphi}_v\right)$
- In SR: $N_v \sim \text{Poisson}\left(\mu_B \vec{\theta} \cdot \vec{\varphi}_v + \mu_S \eta_v\right)$
- Signal distribution $\eta_v$

$$\{\eta_v\}_{v \in \text{SR}} \sim \text{Dir}\left(\frac{1}{\Omega}, \ldots, \frac{1}{\Omega}\right)$$



# Adding signal to the model

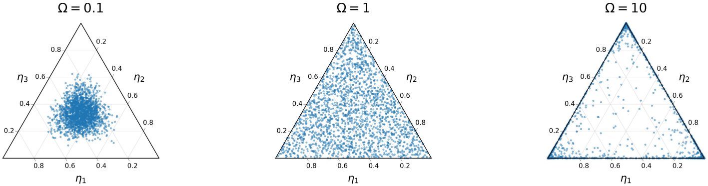



- Signal and background mixture,

$$S \sim \text{Poisson}(\mu_S), \quad B \sim \text{Poisson}(\mu_B)$$

- In CR: $N_v \sim \text{Poisson}\left(\mu_B \vec{\theta} \cdot \vec{\varphi}_v\right)$
- In SR: $N_v \sim \text{Poisson}\left(\mu_B \vec{\theta} \cdot \vec{\varphi}_v + \mu_S \eta_v\right)$
- Signal distribution $\eta_v$

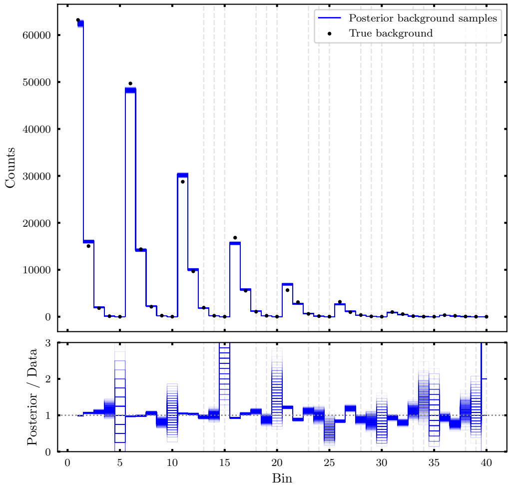
$$\{\eta_v\}_{v \in \text{SR}} \sim \text{Dir}\left(\frac{1}{\Omega}, \ldots, \frac{1}{\Omega}\right)$$

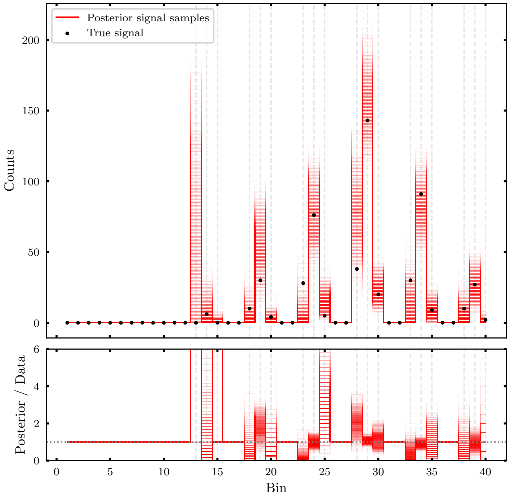



- Given the topics $\{\vec{\varphi}_v\}$, obtain posterior distributions of $\mu_B$, $\mu_S$, $\vec{\theta}$ and $\Omega$

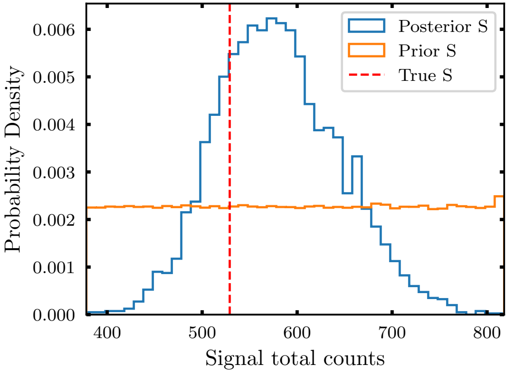
# Preliminary results

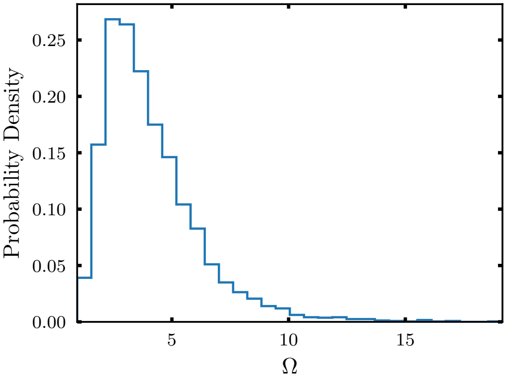
Sample new data from posterior distributions and compare with the training data



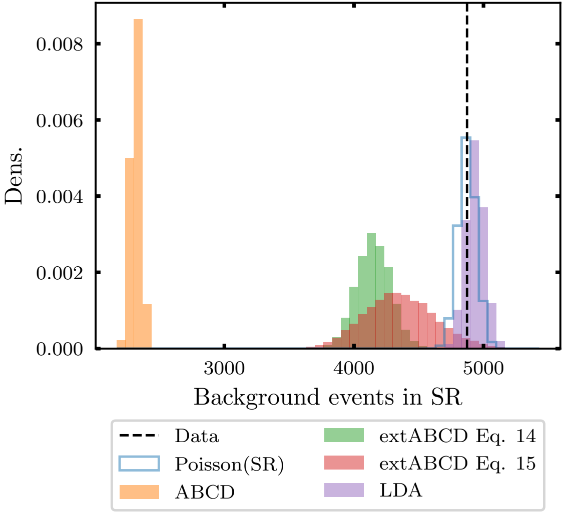



# Preliminary results



# Comparing with ABCDs



# Coverage of signal

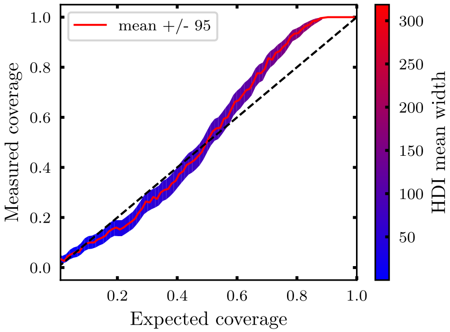
Given the posterior of the number of signal events ($S$), the **smallest credible interval**

$$\mathrm{SCI}_{\alpha\%} = \underset{s_{\min}, s_{\max}}{\arg\min}\,(s_{\max} - s_{\min}) = [s_{\min}, s_{\max}], \qquad \int_{s_{\min}}^{s_{\max}} p(S|\mathbf{d})\,dS = \alpha$$

# Coverage of signal

Given the posterior of the number of signal events ($S$), the **smallest credible interval**

$$\mathrm{SCI}_{\alpha\%} = \underset{s_{\min}, s_{\max}}{\arg\min}\,(s_{\max} - s_{\min}) = [s_{\min}, s_{\max}], \qquad \int_{s_{\min}}^{s_{\max}} p(S|\mathbf{d})\, dS = \alpha$$




- Repeat the process for 500 independent signal samples
- Compute the fraction of times that the $\mathrm{SCI}_{\alpha\%}$ contains the true value of $S$
- Compare the measured coverage with the nominal $\alpha\%$

## Conclusions

- We propose a method based on topic modeling (LDA) for background estimation

# Conclusions

- We propose a method based on topic modeling (LDA) for background estimation

- Test $pp \to t\bar{t}jj$ with $(N_j, N_b)$ as observables
  - Restrict to the case of signals exclusively in a defined SR
    (generalization is $\sim$ straightforward)

# Conclusions

- We propose a method based on topic modeling (LDA) for background estimation

- Test $pp \rightarrow t\bar{t}jj$ with $(N_j, N_b)$ as observables
  - Restrict to the case of signals exclusively in a defined SR
    (generalization is $\sim$ straightforward)

- Extract the latent topic distributions from a set of simulations with different parametrizations
  - Learns correlations between measurements that could be difficult to model
  - Could allow to reduce the impact of inaccurate simulations

# Conclusions

- We propose a method based on topic modeling (LDA) for background estimation

- Test $pp \to t\bar{t}jj$ with $(N_j, N_b)$ as observables
  - Restrict to the case of signals exclusively in a defined SR
    (generalization is $\sim$ straightforward)

- Extract the latent topic distributions from a set of simulations with different parametrizations
  - Learns correlations between measurements that could be difficult to model
  - Could allow to reduce the impact of inaccurate simulations

- Infer the topic mixture in the target background + injected signal
  - Significant improvement in background estimation compared to the traditional ABCD method and some of its extensions

# Conclusions

- We propose a method based on topic modeling (LDA) for background estimation

- Test $pp \to t\bar{t}jj$ with $(N_j, N_b)$ as observables
  - Restrict to the case of signals exclusively in a defined SR
    (generalization is $\sim$ straightforward)

- Extract the latent topic distributions from a set of simulations with different parametrizations
  - Learns correlations between measurements that could be difficult to model
  - Could allow to reduce the impact of inaccurate simulations

- Infer the topic mixture in the target background + injected signal
  - Significant improvement in background estimation compared to the traditional ABCD method and some of its extensions

- Paper is not out yet, but check out 2604.02219 by Álvarez, Benevedes, Szewc, Thaler for a similar application

## Conclusions

- We propose a method based on topic modeling (LDA) for background estimation
- Test $pp \to t\bar{t}jj$ with $(N_j, N_b)$ as observables
  - Restrict to the case of signals exclusively in a defined SR (generalization is $\sim$ straightforward)

**Thank you!**

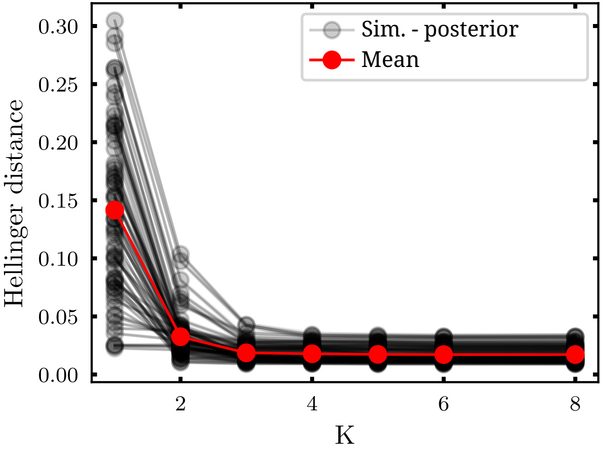
- Infer the topic mixture in the target background + injected signal
  - Significant improvement in background estimation compared to the traditional ABCD method and some of its extensions
- Paper is not out yet, but check out 2604.02219 by Álvarez, Benevedes, Szewc, Thaler for a similar application

# Backup

# Choosing the number of topics

For each simulation, Hellinger distance from posterior to true

$$H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{v=1}^{V}\left(\sqrt{p_v} - \sqrt{q_v}\right)^2}$$



# Best topic mixture allowed by priors