

# **Bayesi modell-összehasonlítás és alkalmazási irányok**

## **Nyelvi adatok feldolgozása**

Borbély Gábor

2020.05.18.

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n, x_j \in X, y_j \in Y$  valamelyen adat.

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n, x_j \in X, y_j \in Y$  valamelyen adat.
- Tegyük fel, hogy egy  $\mathbb{P}(y_j | x_j, \mathbf{w})$  modellt akarok ( $\mathbf{w}$  modell paraméter)

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n, x_j \in X, y_j \in Y$  valamelyen adat.
- Tegyük fel, hogy egy  $\mathbb{P}(y_j | x_j, \mathbf{w})$  modellt akarok ( $\mathbf{w}$  modell paraméter)
  - Nem csak döntést hoz, hanem valószínűségeket is ad.

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n$ ,  $x_j \in X$ ,  $y_j \in Y$  valamelyen adat.
- Tegyük fel, hogy egy  $\mathbb{P}(y_j | x_j, \mathbf{w})$  modellt akarok ( $\mathbf{w}$  modell paraméter)
  - Nem csak döntést hoz, hanem valószínűségeket is ad.
- Vagy a feladat legyen címkézetlen, felügyelet nélküli tanulás (representation learning)

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n, x_j \in X, y_j \in Y$  valamelyen adat.
- Tegyük fel, hogy egy  $\mathbb{P}(y_j | x_j, \mathbf{w})$  modellt akarok ( $\mathbf{w}$  modell paraméter)
  - Nem csak döntést hoz, hanem valószínűségeket is ad.
- Vagy a feladat legyen címkézetlen, felügyelet nélküli tanulás (representation learning)
  - Ekkor a feladatunk megtanulni az adatnak az  $X$  tér feletti eloszlását:

$$\mathbb{P}(x_j | \mathbf{w})$$

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n, x_j \in X, y_j \in Y$  valamelyen adat.
- Tegyük fel, hogy egy  $\mathbb{P}(y_j | x_j, \mathbf{w})$  modellt akarok ( $\mathbf{w}$  modell paraméter)
  - Nem csak döntést hoz, hanem valószínűségeket is ad.
- Vagy a feladat legyen címkézetlen, felügyelet nélküli tanulás (representation learning)
  - Ekkor a feladatunk megtanulni az adatnak az  $X$  tér feletti eloszlását:

$$\mathbb{P}(x_j | \mathbf{w})$$

- a nyelvmodellezés is ilyen (pl. word2vec):

$$\max_{\text{szóvektorok}} \mathbb{P}(\text{szöveg} | \text{szóvektorok})$$

# Gépi tanulás általában

- Legyen  $(x_j, y_j)_{j=1}^n, x_j \in X, y_j \in Y$  valamelyen adat.
- Tegyük fel, hogy egy  $\mathbb{P}(y_j | x_j, \mathbf{w})$  modellt akarok ( $\mathbf{w}$  modell paraméter)
  - Nem csak döntést hoz, hanem valószínűségeket is ad.
- Vagy a feladat legyen címkézetlen, felügyelet nélküli tanulás (representation learning)
  - Ekkor a feladatunk megtanulni az adatnak az  $X$  tér feletti eloszlását:

$$\mathbb{P}(x_j | \mathbf{w})$$

- a nyelvmodellezés is ilyen (pl. word2vec):

$$\max_{\text{szóvektorok}} \mathbb{P}(\text{szöveg} | \text{szóvektorok})$$

- A most következő módszer használható címkézett (felügyelt) tanításhoz is.

# Representation Learning

- GAN, WGAN

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?
  - Gondolhatnánk, hogy ha jobban illeszkedik az adatra, akkor jobb.

$$\text{dist} (\mathbb{P}(x_j | \mathbf{w}_i), \mathbb{P}(x_j)) \rightarrow \min$$

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?
  - Gondolhatnánk, hogy ha jobban illeszkedik az adatra, akkor jobb.

$$\text{dist} (\mathbb{P}(x_j | \mathbf{w}_i), \mathbb{P}(x_j)) \rightarrow \min$$

- De ekkor mindenkor a memorizálás a legjobb.

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?
  - Gondolhatnánk, hogy ha jobban illeszkedik az adatra, akkor jobb.

$$\text{dist} (\mathbb{P}(x_j | \mathbf{w}_i), \mathbb{P}(x_j)) \rightarrow \min$$

- De ekkor mindenkor a memorizálás a legjobb.
  - Érdekesség: egy neurális háló (néha) nagyobb, mint az az adat, amin feltanították.

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?
  - Gondolhatnánk, hogy ha jobban illeszkedik az adatra, akkor jobb.

$$\text{dist} (\mathbb{P}(x_j | \mathbf{w}_i), \mathbb{P}(x_j)) \rightarrow \min$$

- De ekkor mindenkor a memorizálás a legjobb.
  - Érdekesség: egy neurális háló (néha) nagyobb, mint az az adat, amin feltanították.
- Például nyelvmodellezésben ha túl nagy  $n$ -gramokat veszünk

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?
  - Gondolhatnánk, hogy ha jobban illeszkedik az adatra, akkor jobb.

$$\text{dist} (\mathbb{P}(x_j | \mathbf{w}_i), \mathbb{P}(x_j)) \rightarrow \min$$

- De ekkor mindenkor a memorizálás a legjobb.
  - Érdekesség: egy neurális háló (néha) nagyobb, mint az az adat, amin feltanították.
  - Például nyelvmodellezésben ha túl nagy  $n$ -gramokat vesznek
- Regularizáció

# Representation Learning

- GAN, WGAN
- LM, a szöveg perplexitásának minimalizálása (a modell paraméterekben)
- Mi van, ha több modell is van?
- Legyen  $\mathbf{w}_i \in \mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  több tér, melyek mindegyikében még  $\mathbf{w}_i$ -t is ki kell választanom.
- Melyik a jobb?
  - Gondolhatnánk, hogy ha jobban illeszkedik az adatra, akkor jobb.

$$\text{dist} (\mathbb{P}(x_j | \mathbf{w}_i), \mathbb{P}(x_j)) \rightarrow \min$$

- De ekkor mindenkor a memorizálás a legjobb.
  - Érdekesség: egy neurális háló (néha) nagyobb, mint az az adat, amin feltanították.
  - Például nyelvmodellezésben ha túl nagy  $n$ -gramokat vesznek
- Regularizáció
- Occam's razor: lehet rosszabb az illeszkedés, ha kisebb a modell.

# Bayesi modell-összehasonlítás – I.

- Legyen  $D \subseteq X$  adat, egy adatpont többször is szerepelhet.

# Bayesi modell-összehasonlítás – I.

- Legyen  $D \subseteq X$  adat, egy adatpont többször is szerepelhet.
  - multiset, multiplicitás:  $(x_j, n_j)$  párok, ahol  $n_j \in \mathbb{N}$ .

# Bayesi modell-összehasonlítás – I.

- Legyen  $D \subseteq X$  adat, egy adatpont többször is szerepelhet.
  - multiset, multiplicitás:  $(x_j, n_j)$  párok, ahol  $n_j \in \mathbb{N}$ .
- Adott  $\mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  terek és mindegyikben egy valószínűségi modell:

$$\mathbb{P}(\bullet \mid \mathbf{w}_i) \quad \mathbf{w}_i \in \mathcal{H}_i$$

Ezek közül kell választanunk.

# Bayesi modell-összehasonlítás – I.

- Legyen  $D \subseteq X$  adat, egy adatpont többször is szerepelhet.
  - multiset, multiplicitás:  $(x_j, n_j)$  párok, ahol  $n_j \in \mathbb{N}$ .
- Adott  $\mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  terek és mindegyikben egy valószínűségi modell:

$$\mathbb{P}(\bullet \mid \mathbf{w}_i) \quad \mathbf{w}_i \in \mathcal{H}_i$$

Ezek közül kell választanunk.

- Bayes téTEL:

$$\text{"evidence"} = \mathbb{P}(\mathcal{H}_i \mid D) = \frac{\mathbb{P}(D \mid \mathcal{H}_i) \cdot \mathbb{P}(\mathcal{H}_i)}{\mathbb{P}(D)}$$

# Bayesi modell-összehasonlítás – I.

- Legyen  $D \subseteq X$  adat, egy adatpont többször is szerepelhet.
  - multiset, multiplicitás:  $(x_j, n_j)$  párok, ahol  $n_j \in \mathbb{N}$ .
- Adott  $\mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  terek és mindegyikben egy valószínűségi modell:

$$\mathbb{P}(\bullet \mid \mathbf{w}_i) \quad \mathbf{w}_i \in \mathcal{H}_i$$

Ezek közül kell választanunk.

- Bayes téTEL:

$$\text{"evidence"} = \mathbb{P}(\mathcal{H}_i \mid D) = \frac{\mathbb{P}(D \mid \mathcal{H}_i) \cdot \mathbb{P}(\mathcal{H}_i)}{\mathbb{P}(D)}$$

- Ha  $\mathbb{P}(\mathcal{H}_i)$  konstans (nincs a priori preferált modell)

$$\mathbb{P}(\mathcal{H}_i \mid D) \propto \mathbb{P}(D \mid \mathcal{H}_i) = \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) \cdot \mathbb{P}(\mathbf{w}_i \mid \mathcal{H}_i) d\mathbf{w}_i$$

# Bayesi modell-összehasonlítás – I.

- Legyen  $D \subseteq X$  adat, egy adatpont többször is szerepelhet.
  - multiset, multiplicitás:  $(x_j, n_j)$  párok, ahol  $n_j \in \mathbb{N}$ .
- Adott  $\mathcal{H}_i \subseteq \mathbb{R}^{d_i}$  terek és mindegyikben egy valószínűségi modell:

$$\mathbb{P}(\bullet \mid \mathbf{w}_i) \quad \mathbf{w}_i \in \mathcal{H}_i$$

Ezek közül kell választanunk.

- Bayes téTEL:

$$\text{"evidence"} = \mathbb{P}(\mathcal{H}_i \mid D) = \frac{\mathbb{P}(D \mid \mathcal{H}_i) \cdot \mathbb{P}(\mathcal{H}_i)}{\mathbb{P}(D)}$$

- Ha  $\mathbb{P}(\mathcal{H}_i)$  konstans (nincs a priori preferált modell)

$$\mathbb{P}(\mathcal{H}_i \mid D) \propto \mathbb{P}(D \mid \mathcal{H}_i) = \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) \cdot \mathbb{P}(\mathbf{w}_i \mid \mathcal{H}_i) d\mathbf{w}_i$$

- Ha  $\mathbb{P}(\mathbf{w}_i \mid \mathcal{H}_i)$  konstans (nincs a priori preferált paraméter érték)

$$\max_i \frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$$

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket
- Lagrange integrál közelítő módszer

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket
- Lagrange integrál közelítő módszer
- az alábbiak mind közrejátszanak a döntésben:

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket
- Lagrange integrál közelítő módszer
- az alábbiak mind közrejátszanak a döntésben:
  - az adat mérete:  $n$

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket
- Lagrange integrál közelítő módszer
- az alábbiak mind közrejátszanak a döntésben:
  - az adat mérete:  $n$
  - a modell mérete:  $d_i (\text{Vol}(\mathcal{H}_i))$

## Bayesi modell-összehasonlítás – II.

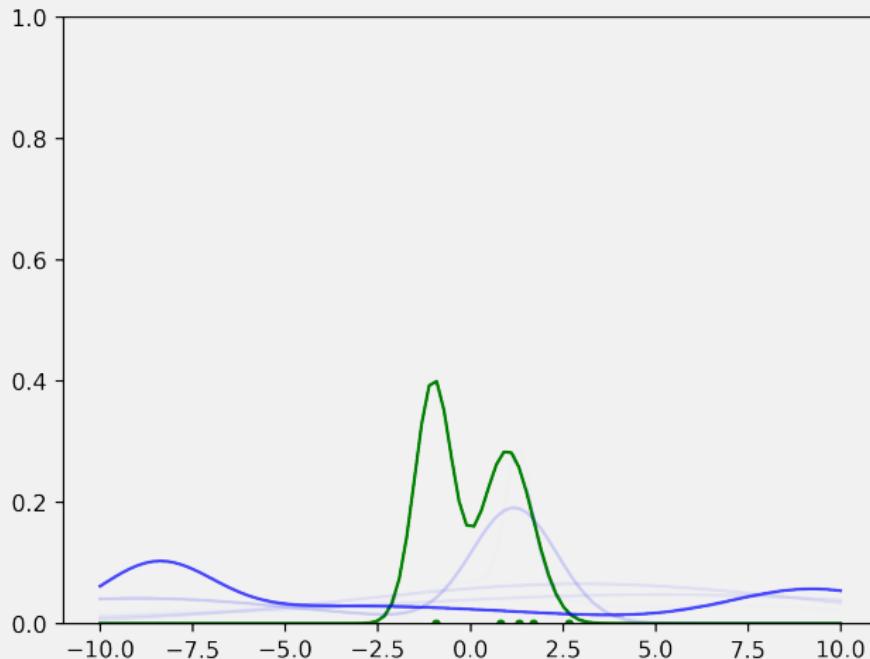
- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket
- Lagrange integrál közelítő módszer
- az alábbiak mind közrejátszanak a döntésben:
  - az adat mérete:  $n$
  - a modell mérete:  $d_i (\text{Vol}(\mathcal{H}_i))$
  - az illesztés (relatív entrópia, KL divergencia)

## Bayesi modell-összehasonlítás – II.

- $\frac{1}{\text{Vol}(\mathcal{H}_i)} \int_{\mathcal{H}_i} \mathbb{P}(D \mid \mathbf{w}_i, \mathcal{H}_i) d\mathbf{w}_i$
- (MacKay, 2003)
- statisztikus fizikában számolnak ilyeneket
- Lagrange integrál közelítő módszer
- az alábbiak mind közrejátszanak a döntésben:
  - az adat mérete:  $n$
  - a modell mérete:  $d_i (\text{Vol}(\mathcal{H}_i))$
  - az illesztés (relatív entrópia, KL divergencia)
- szükség van az illesztés (célfüggvény) Hesse mátrixára

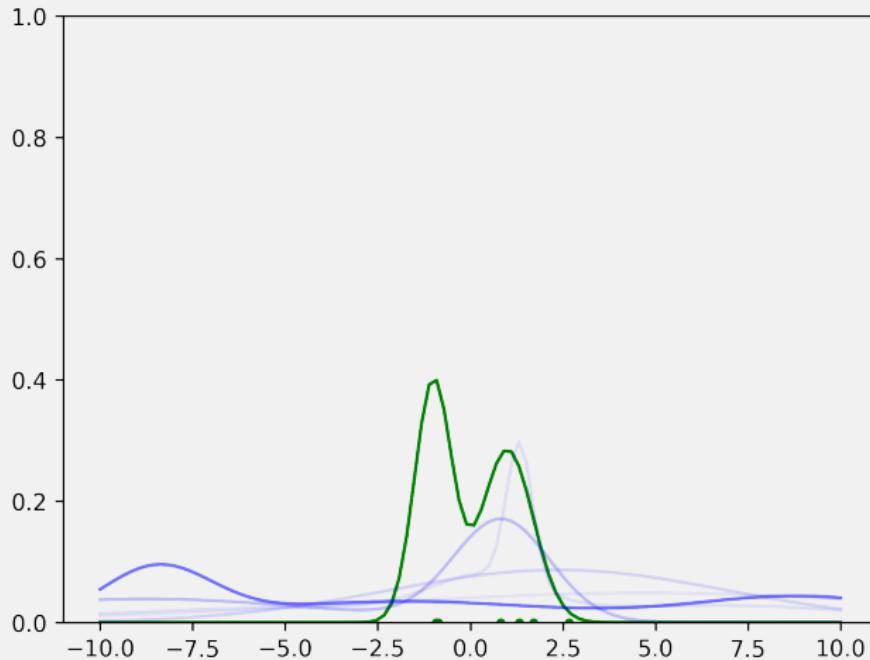
# Bayesi modell-összehasonlítás – demo

$n=5, k=3$



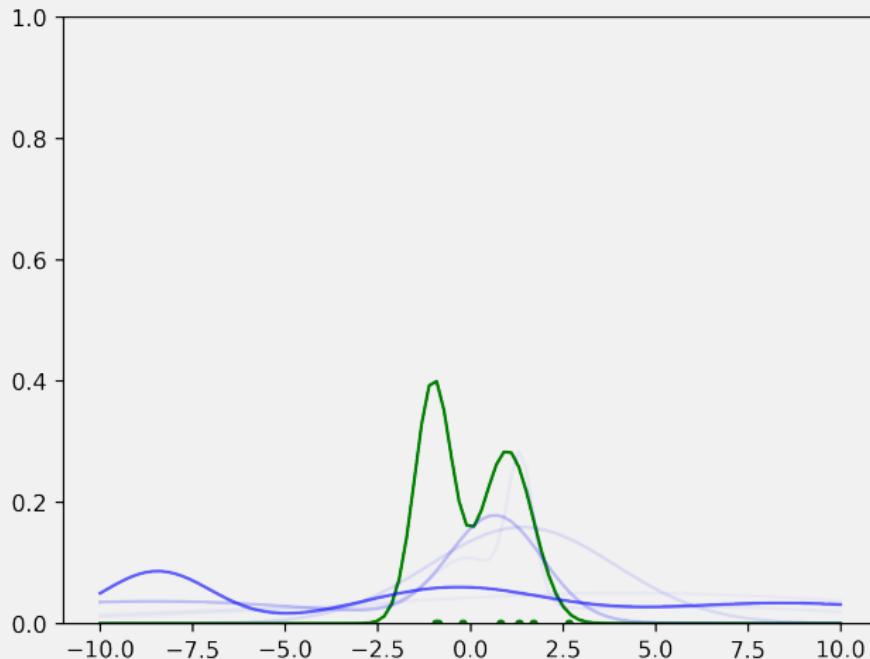
# Bayesi modell-összehasonlítás – demo

$n=6, k=3$



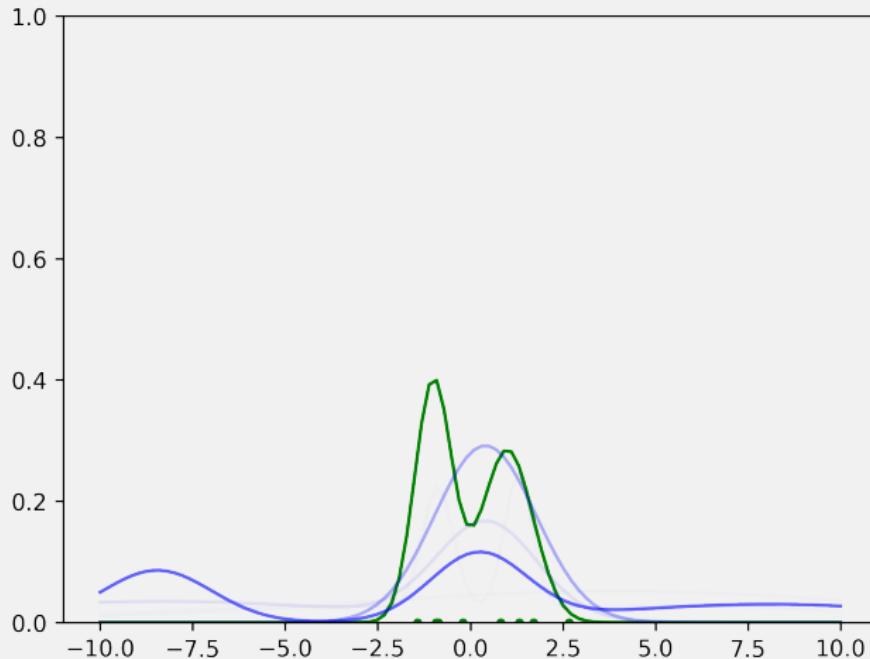
# Bayesi modell-összehasonlítás – demo

$n=7, k=3$



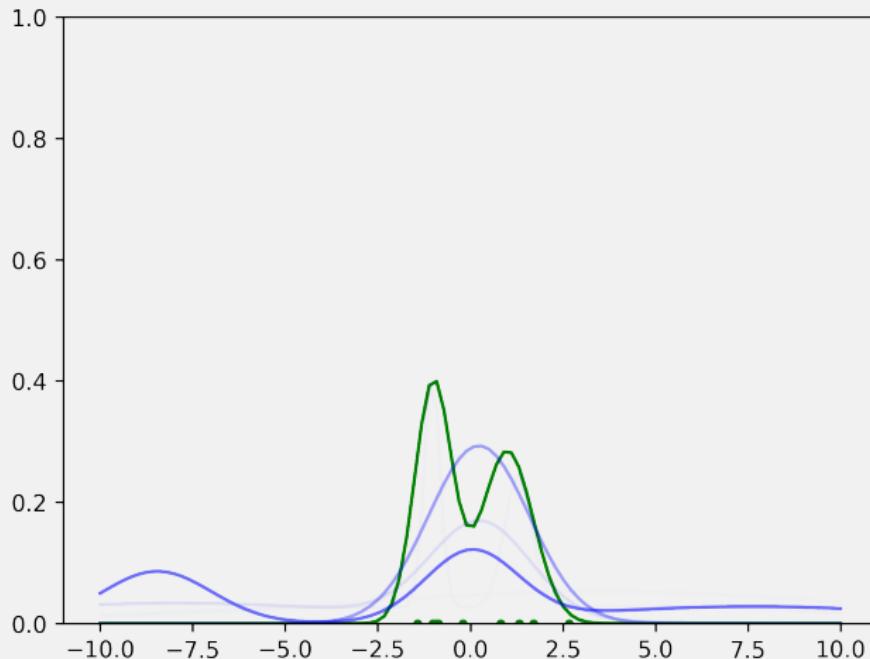
# Bayesi modell-összehasonlítás – demo

$n=8, k=3$



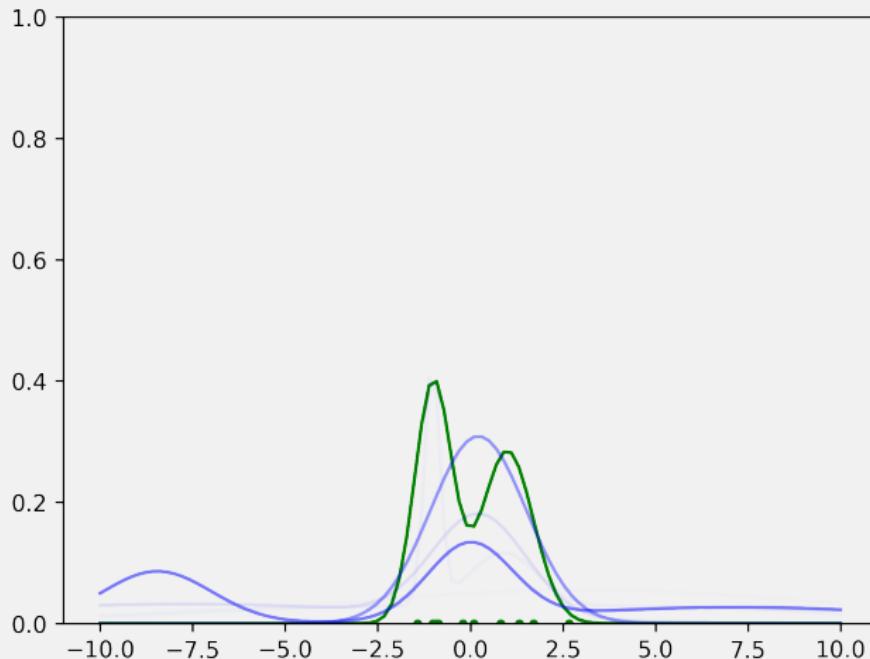
# Bayesi modell-összehasonlítás – demo

$n=9, k=3$



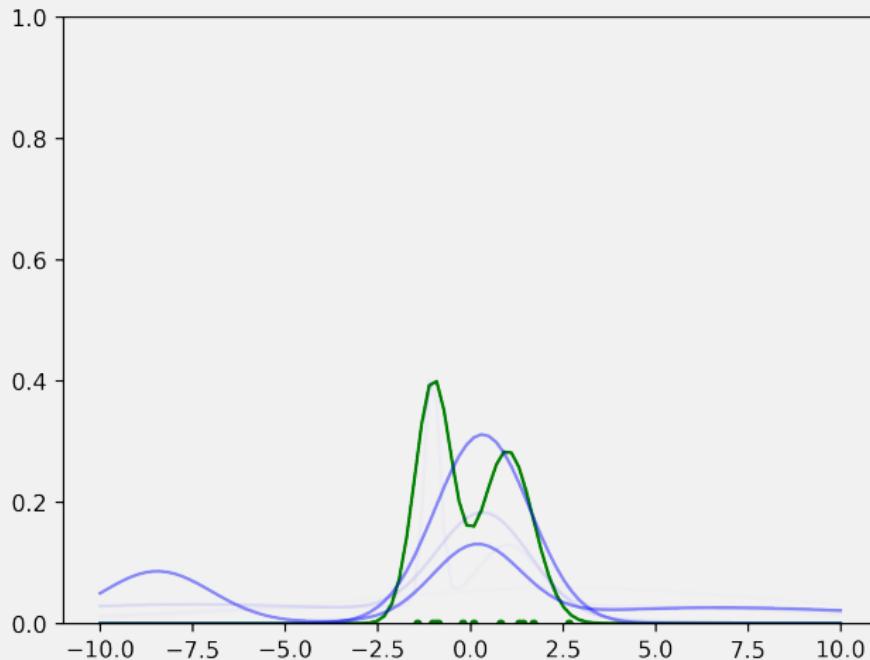
# Bayesi modell-összehasonlítás – demo

n=10, k=3



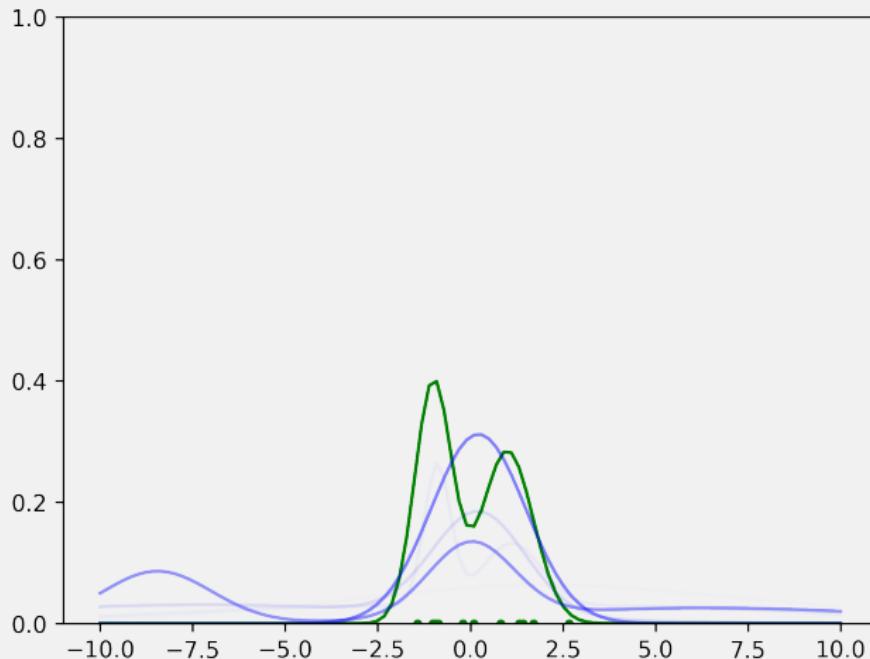
# Bayesi modell-összehasonlítás – demo

n=11, k=1



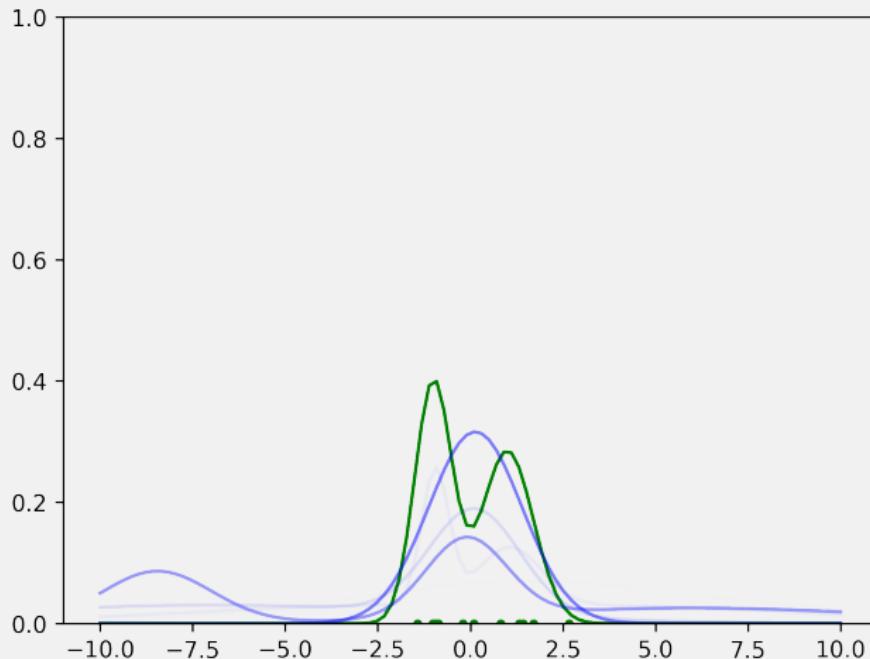
# Bayesi modell-összehasonlítás – demo

$n=12, k=1$



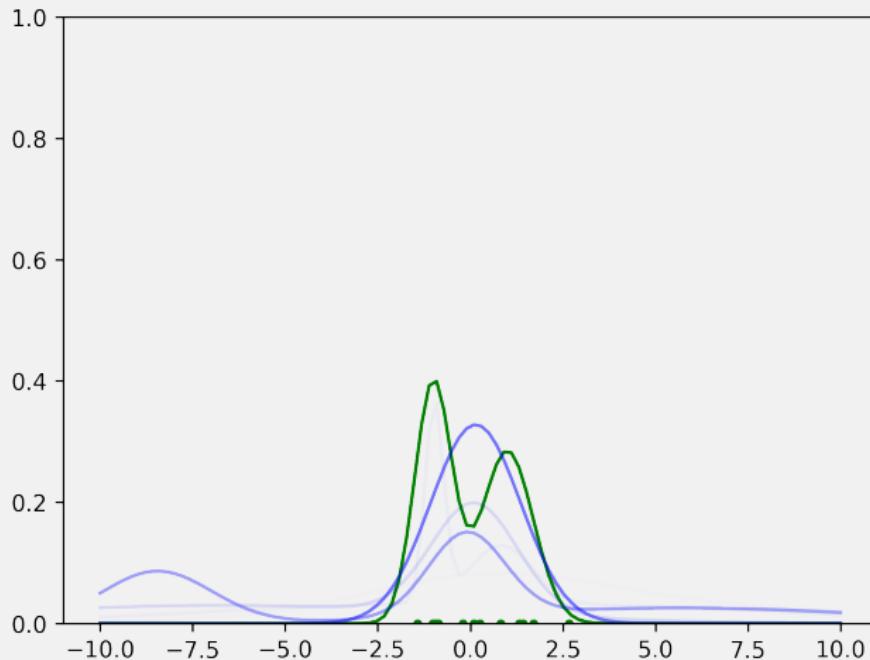
# Bayesi modell-összehasonlítás – demo

$n=13, k=1$



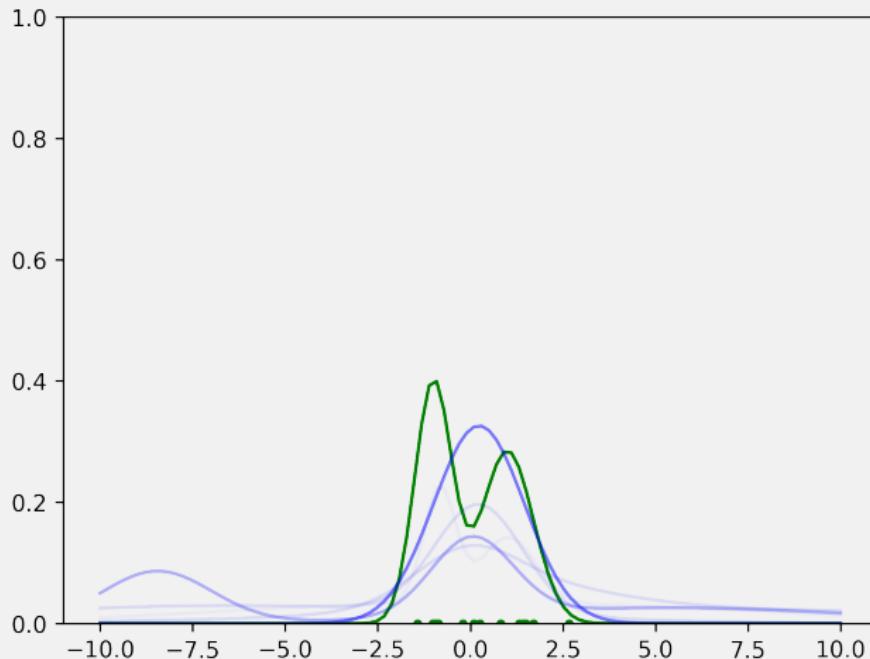
# Bayesi modell-összehasonlítás – demo

$n=14, k=1$



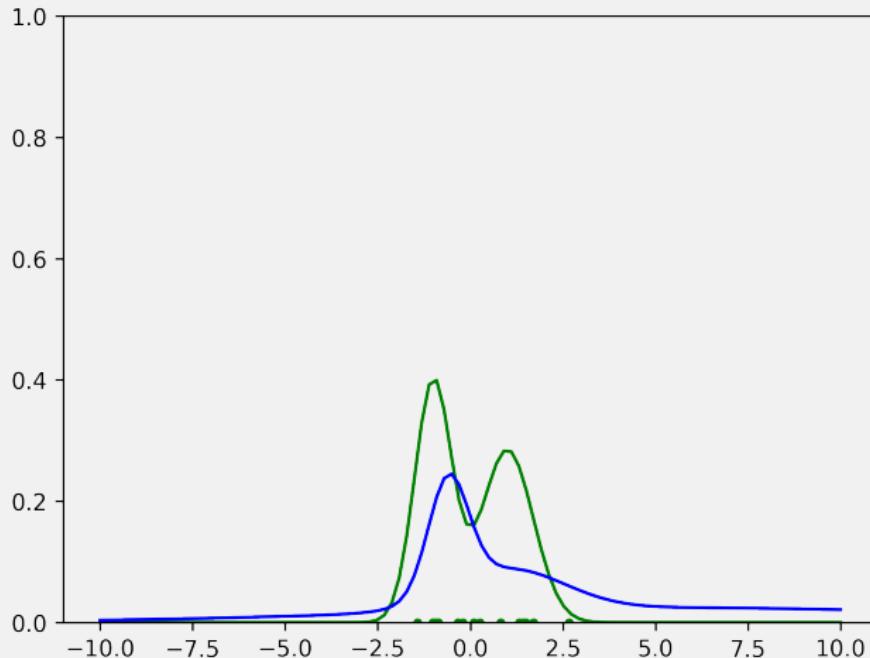
# Bayesi modell-összehasonlítás – demo

$n=15, k=1$



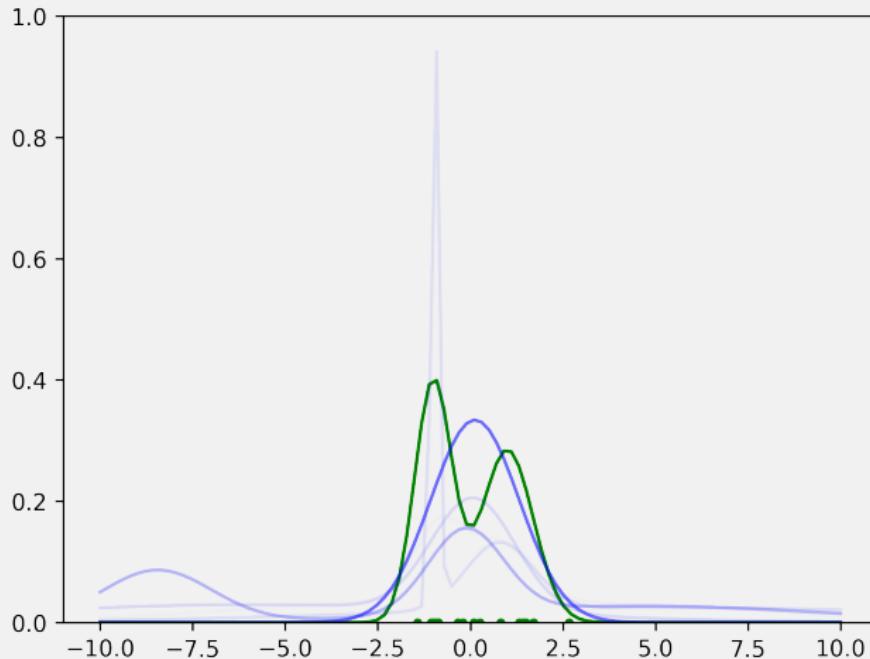
# Bayesi modell-összehasonlítás – demo

$n=16, k=4$



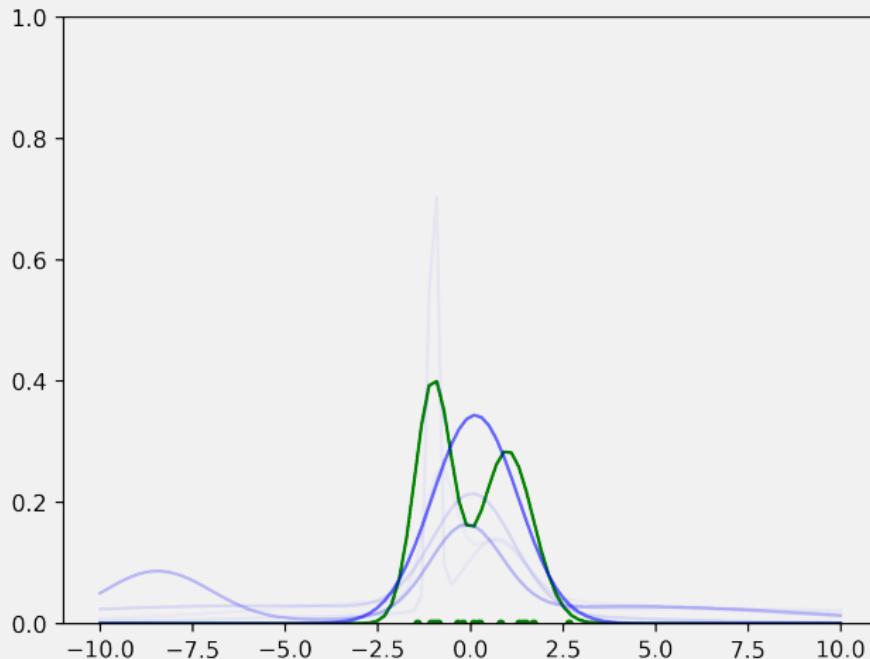
# Bayesi modell-összehasonlítás – demo

$n=17, k=1$



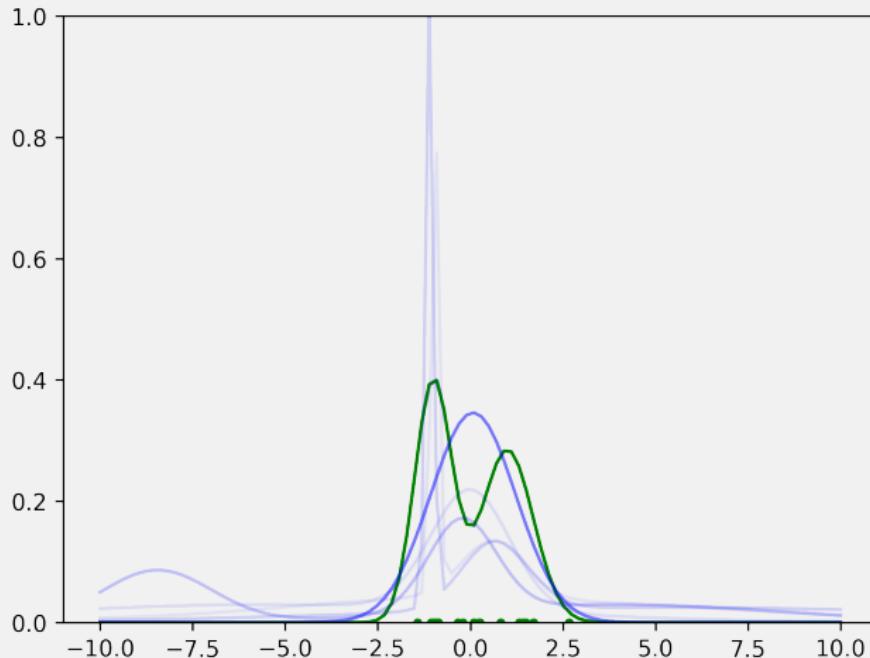
# Bayesi modell-összehasonlítás – demo

$n=18, k=1$



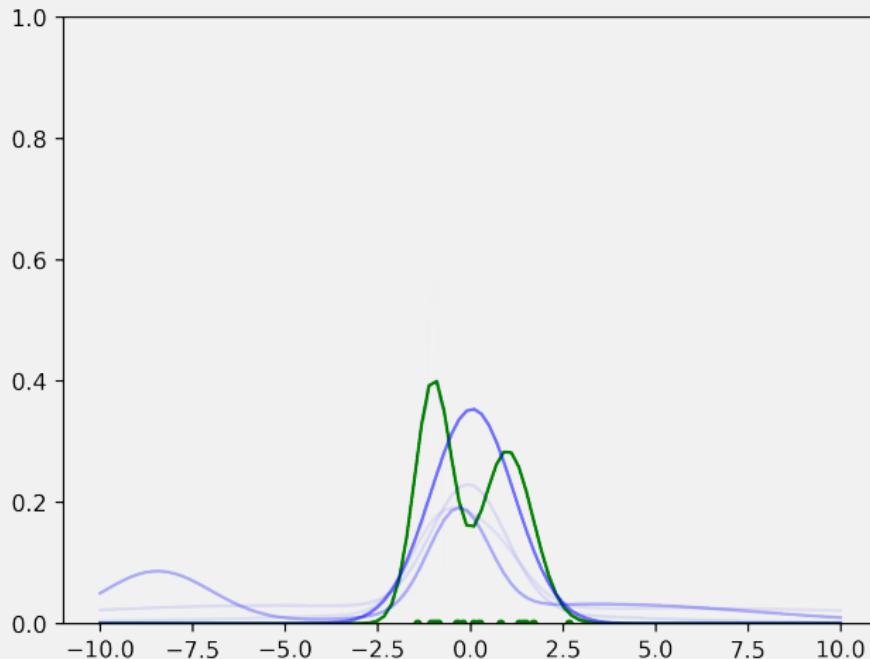
# Bayesi modell-összehasonlítás – demo

$n=19, k=1$



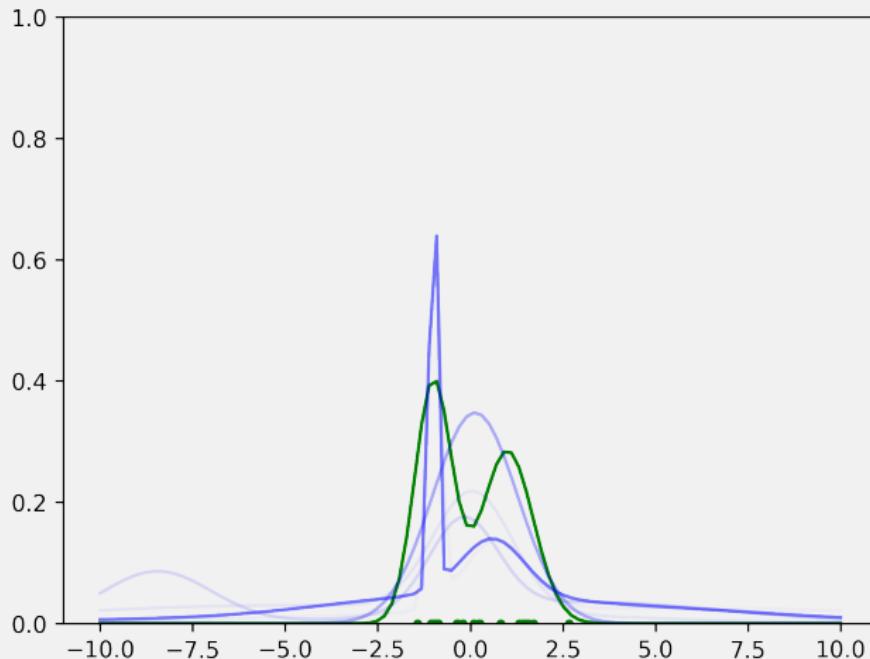
# Bayesi modell-összehasonlítás – demo

$n=20, k=1$



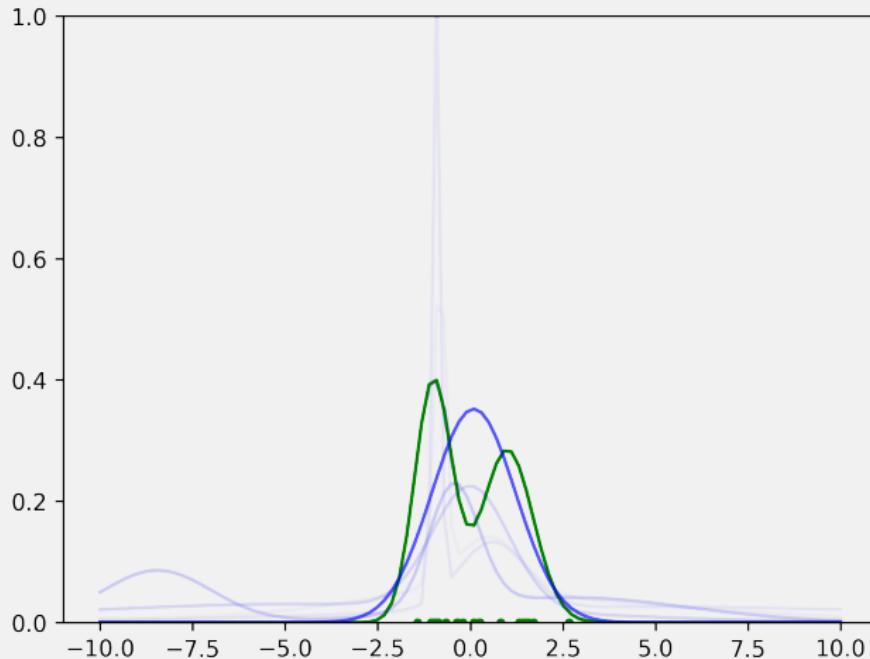
# Bayesi modell-összehasonlítás – demo

n=21, k=5



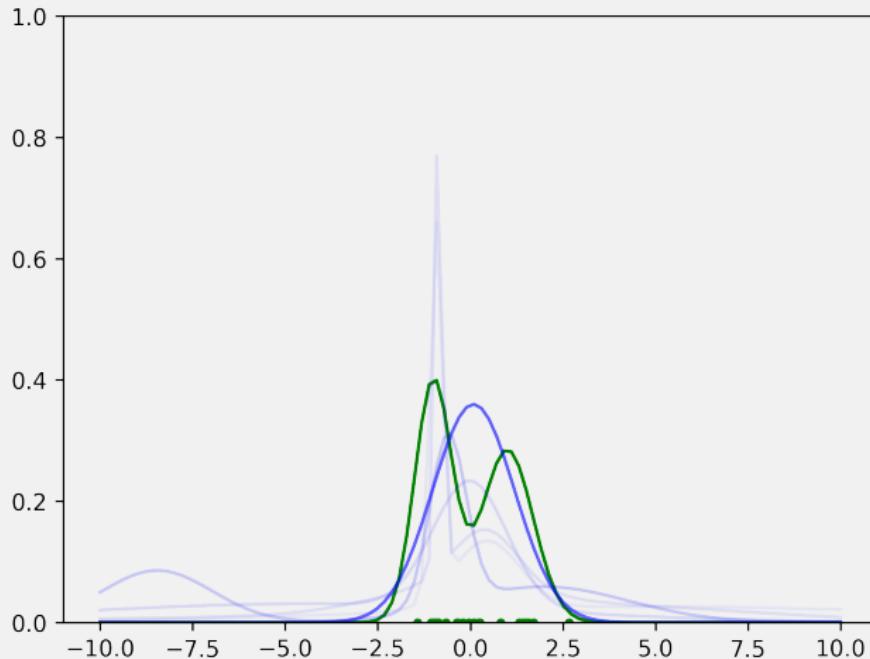
# Bayesi modell-összehasonlítás – demo

n=22, k=1



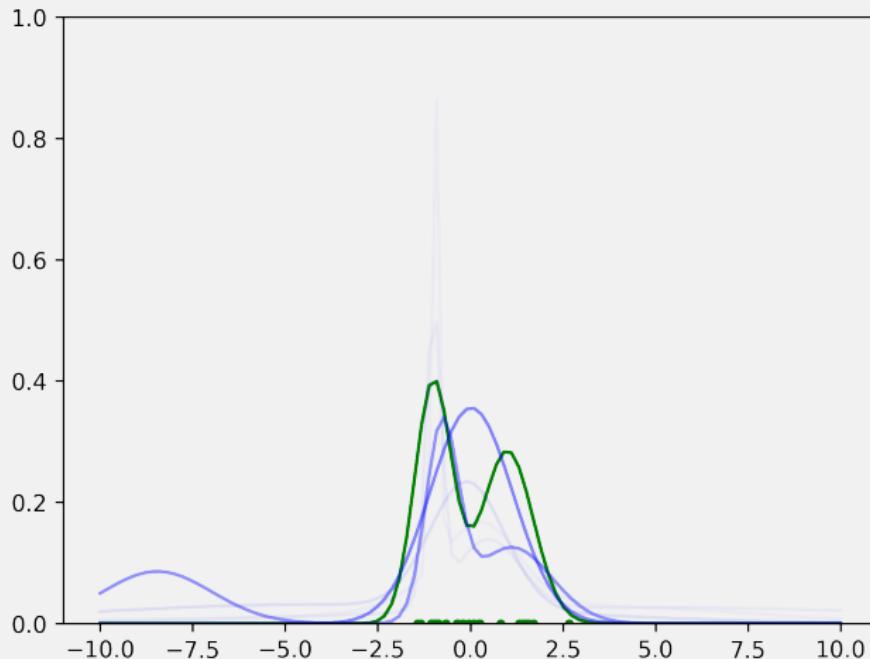
# Bayesi modell-összehasonlítás – demo

$n=23, k=1$



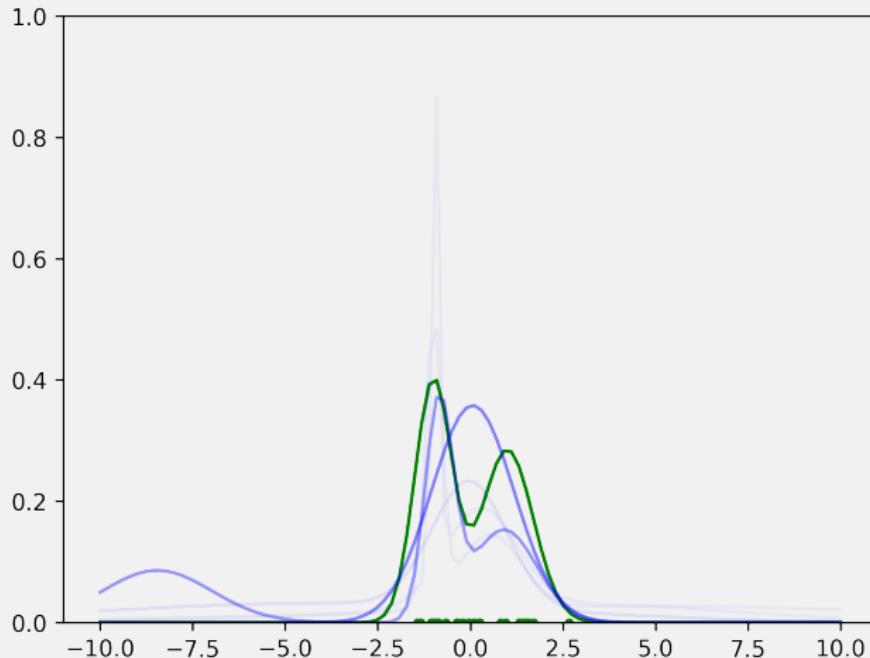
# Bayesi modell-összehasonlítás – demo

$n=24, k=1$



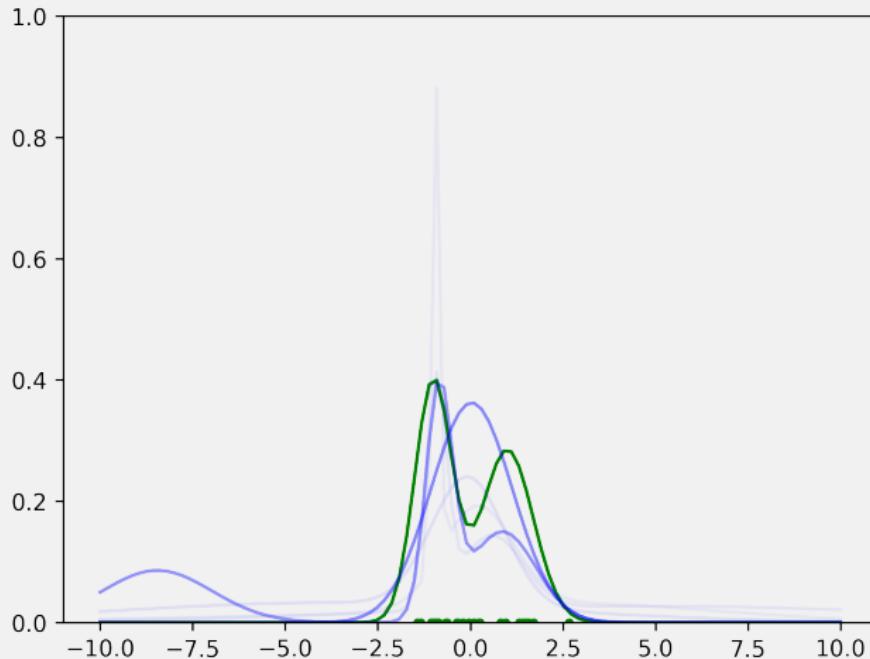
# Bayesi modell-összehasonlítás – demo

$n=25, k=1$



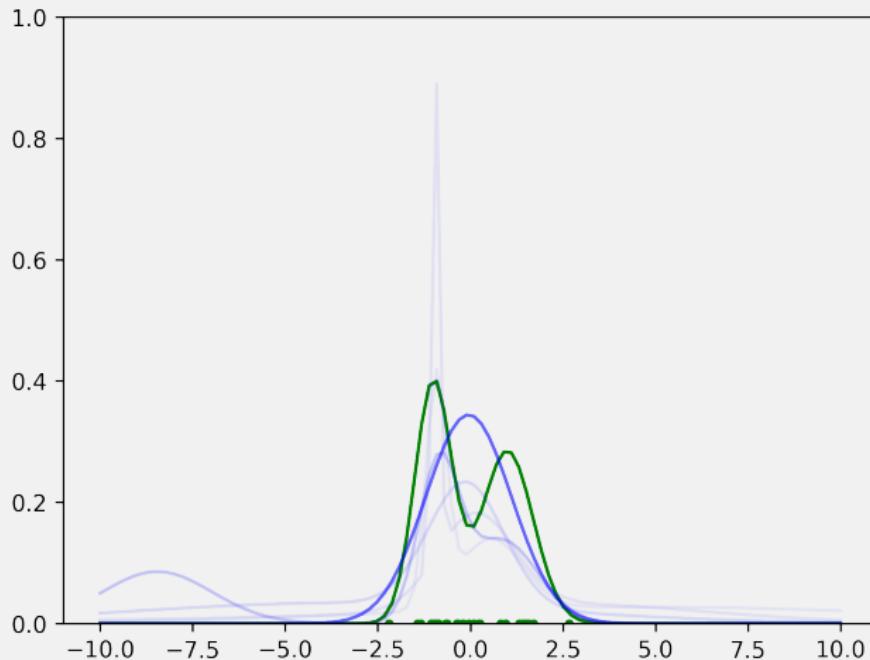
# Bayesi modell-összehasonlítás – demo

$n=26, k=1$



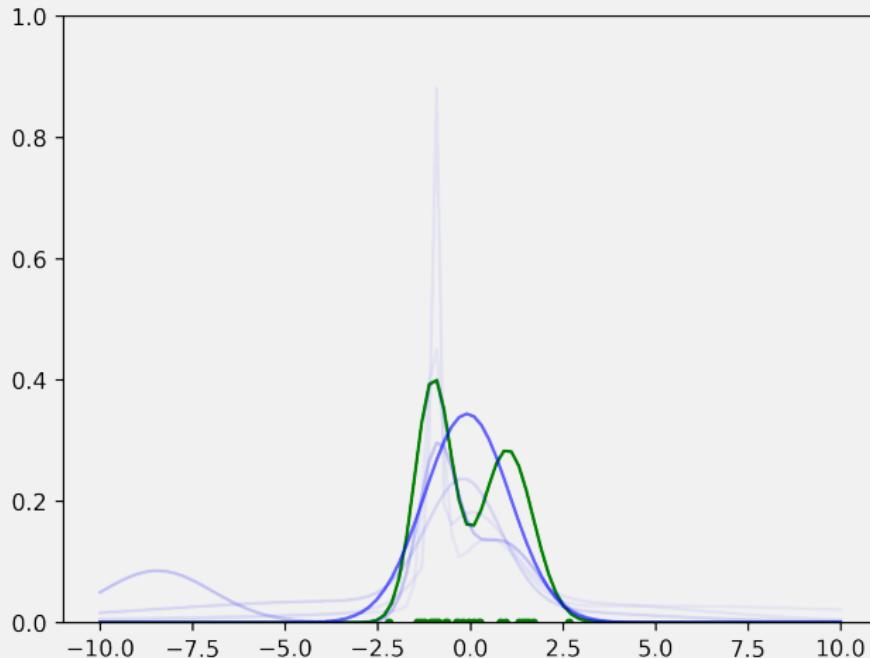
# Bayesi modell-összehasonlítás – demo

n=27, k=1



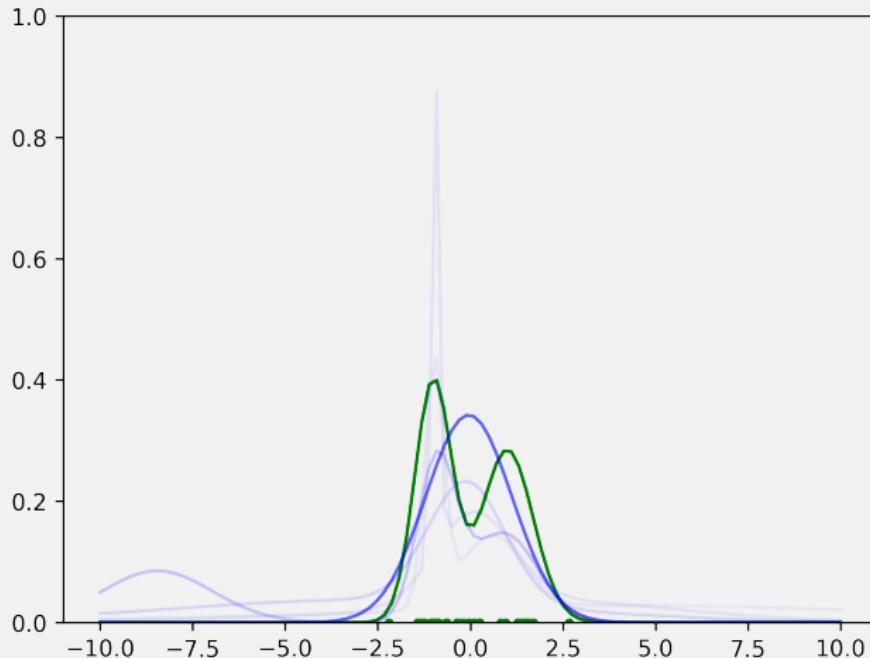
# Bayesi modell-összehasonlítás – demo

$n=28, k=1$



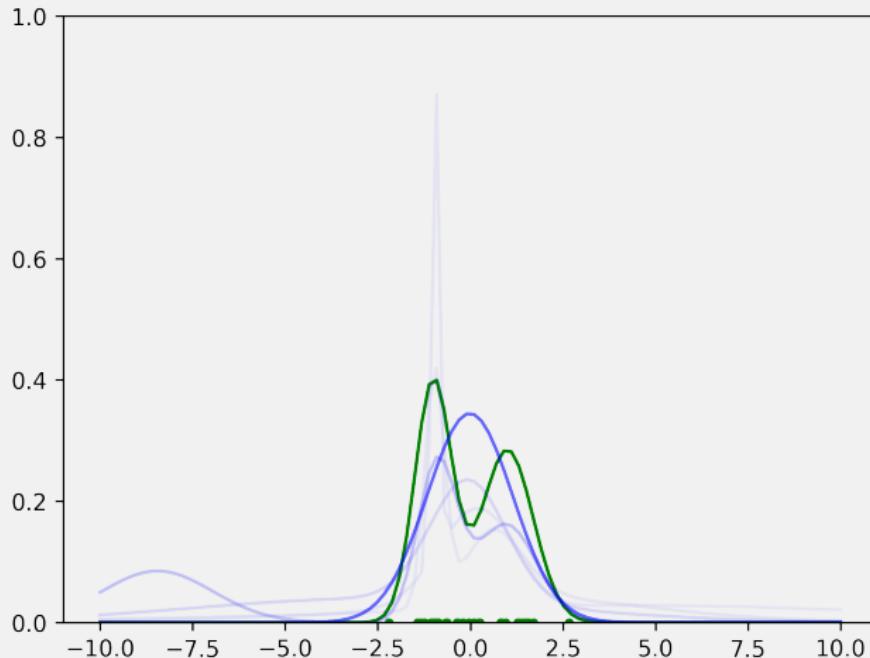
# Bayesi modell-összehasonlítás – demo

$n=29, k=1$



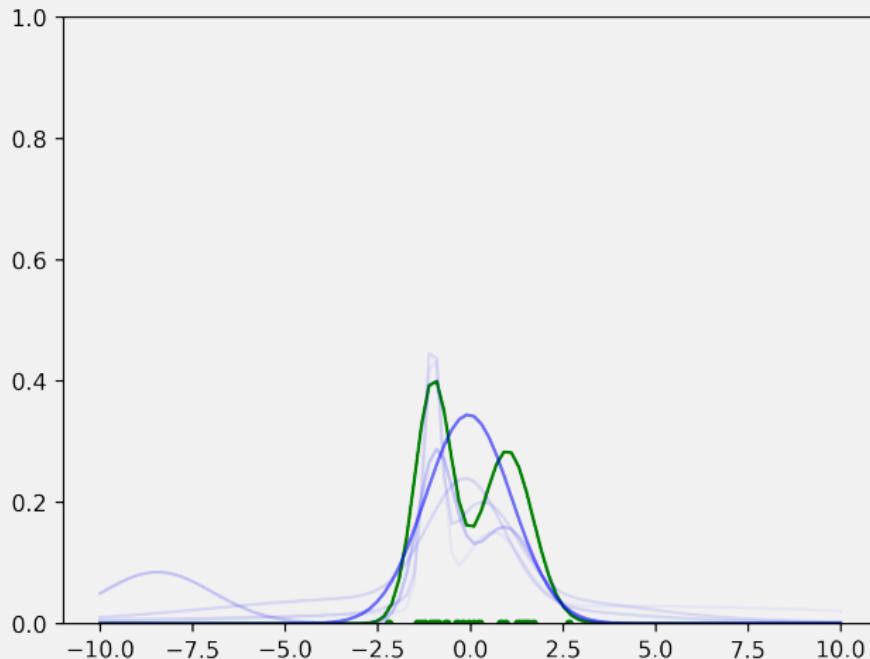
# Bayesi modell-összehasonlítás – demo

n=30, k=1



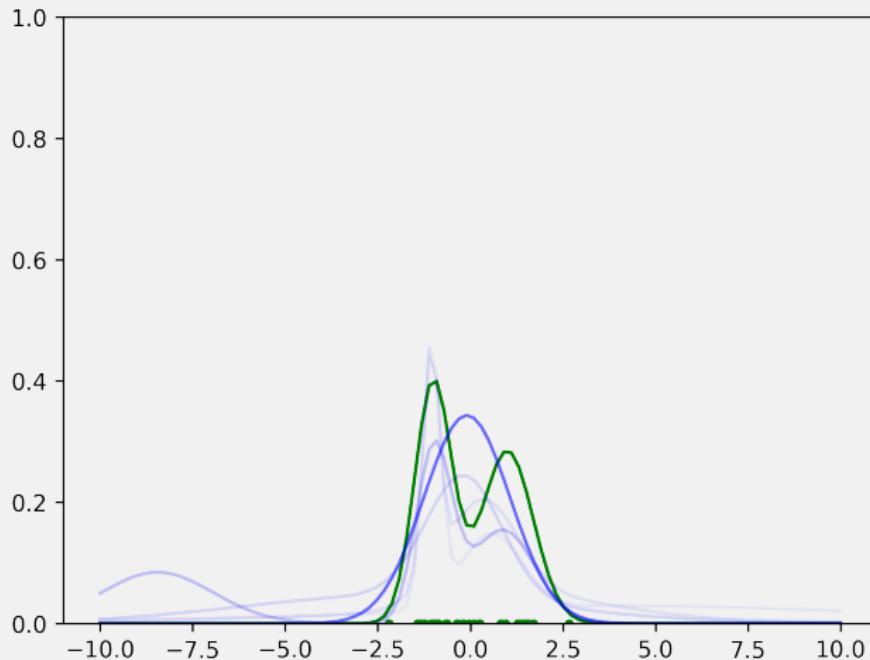
# Bayesi modell-összehasonlítás – demo

n=31, k=1



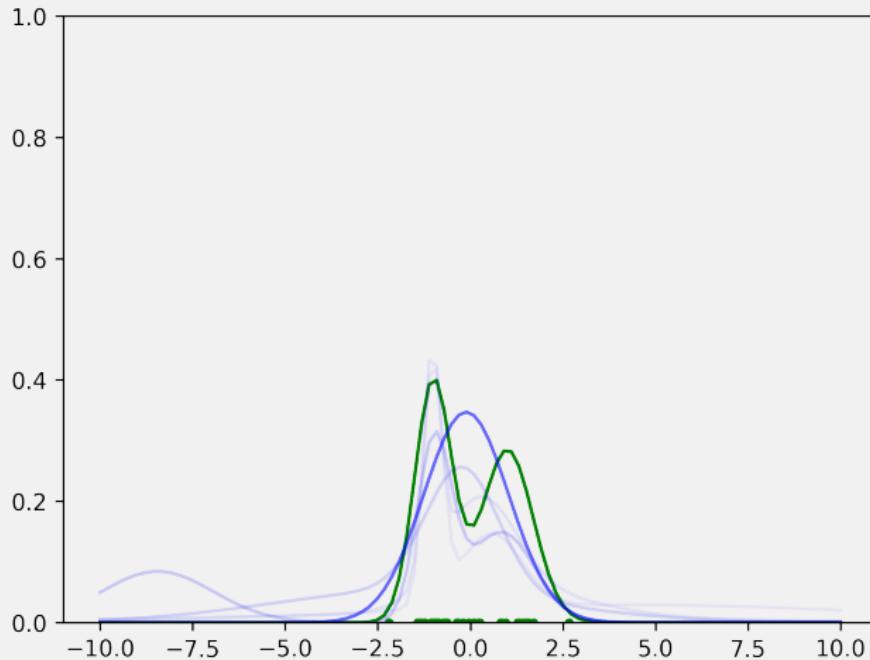
# Bayesi modell-összehasonlítás – demo

n=32, k=1



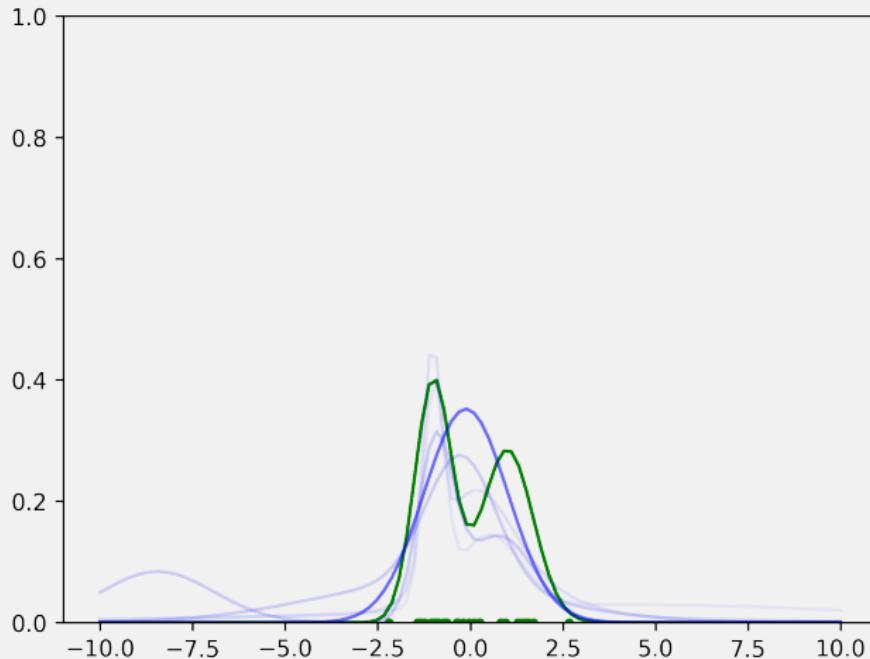
# Bayesi modell-összehasonlítás – demo

n=33, k=1



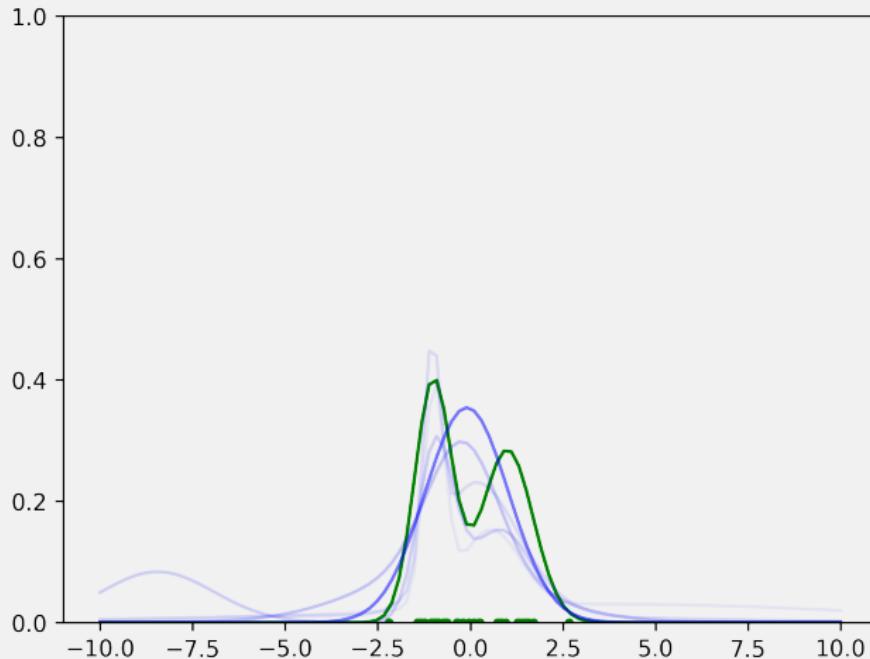
# Bayesi modell-összehasonlítás – demo

n=34, k=1



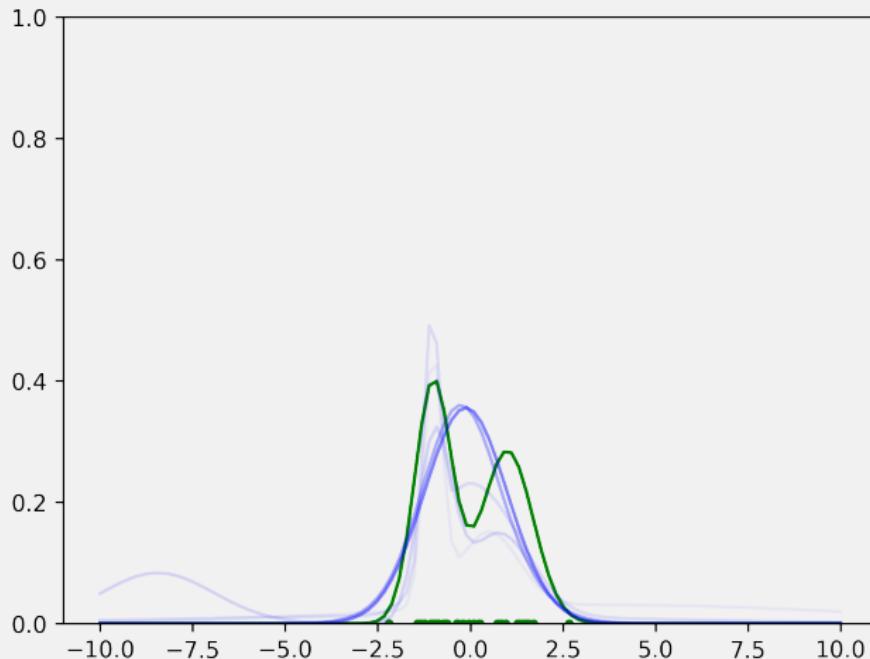
# Bayesi modell-összehasonlítás – demo

n=35, k=1



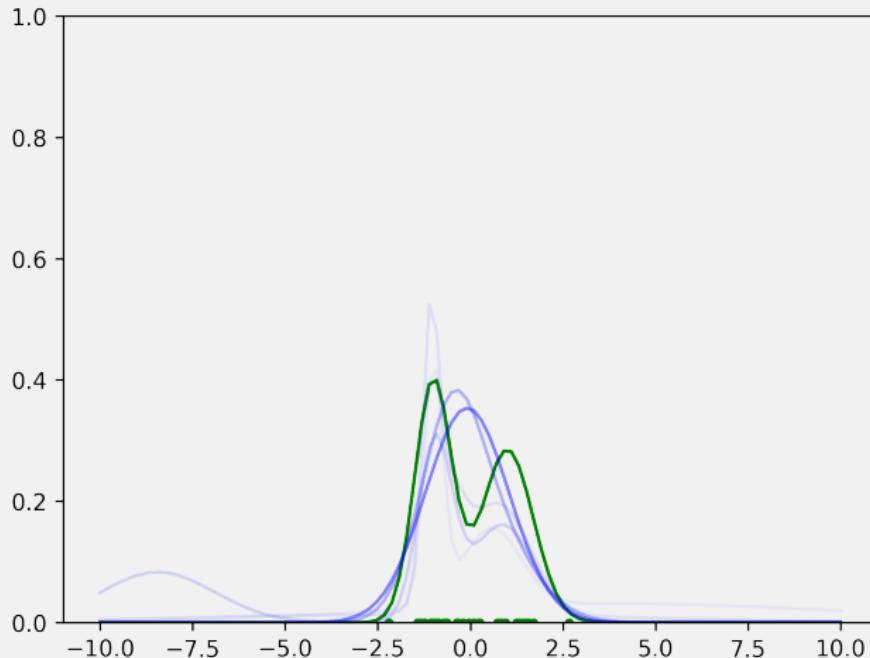
# Bayesi modell-összehasonlítás – demo

n=36, k=1



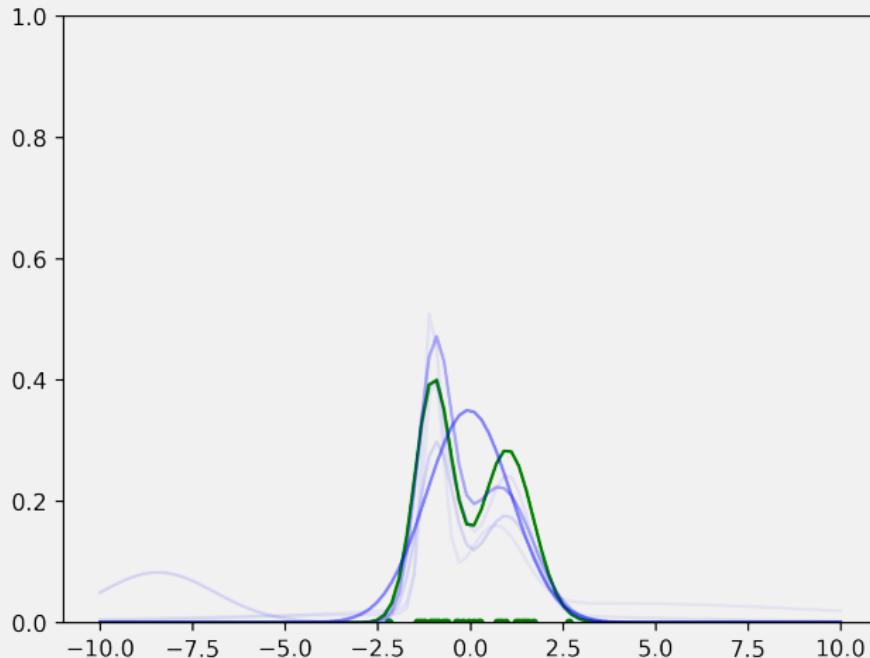
# Bayesi modell-összehasonlítás – demo

n=37, k=1



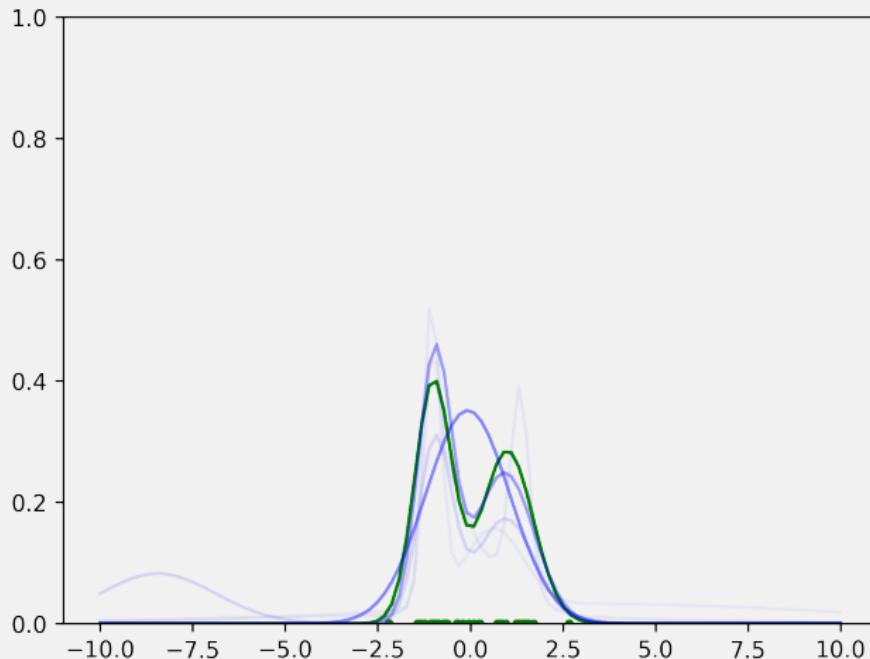
# Bayesi modell-összehasonlítás – demo

n=38, k=1



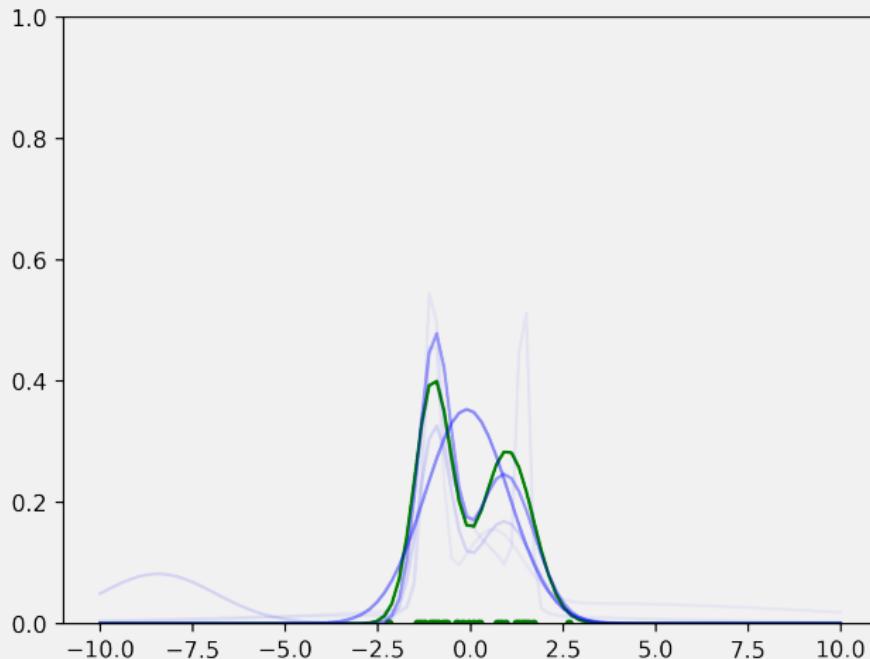
# Bayesi modell-összehasonlítás – demo

n=39, k=1



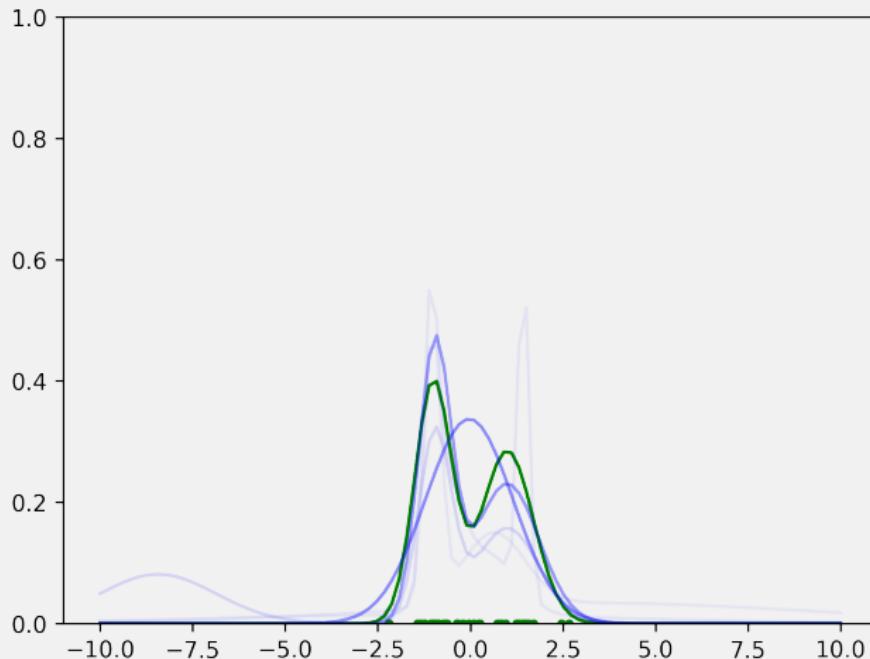
# Bayesi modell-összehasonlítás – demo

$n=40, k=1$



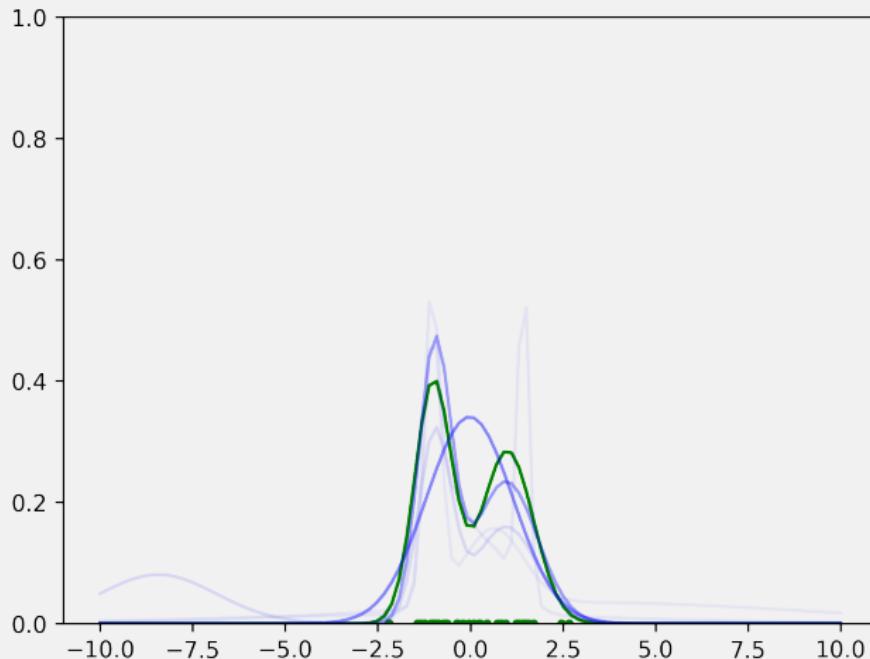
# Bayesi modell-összehasonlítás – demo

$n=41, k=1$



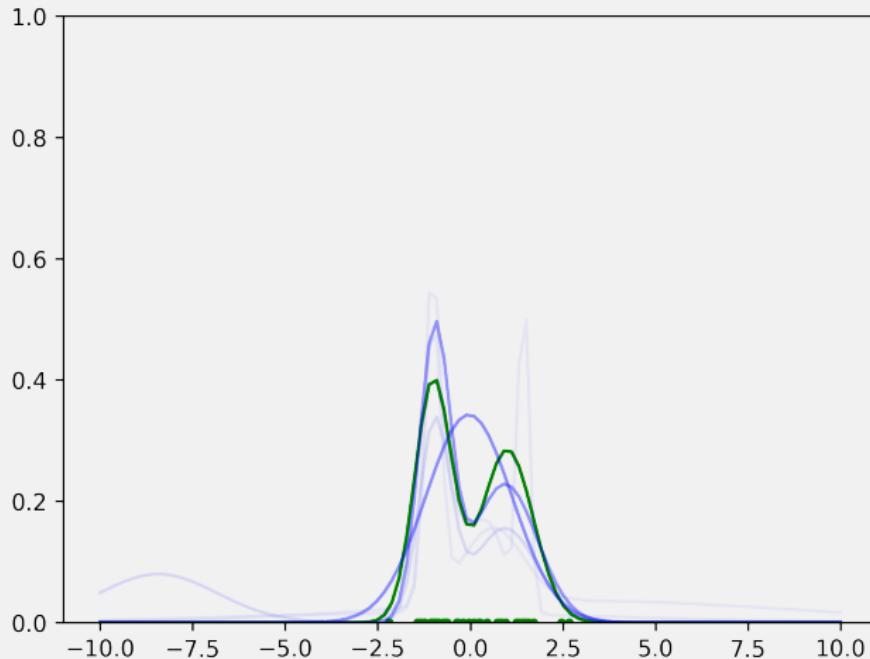
# Bayesi modell-összehasonlítás – demo

$n=42, k=1$



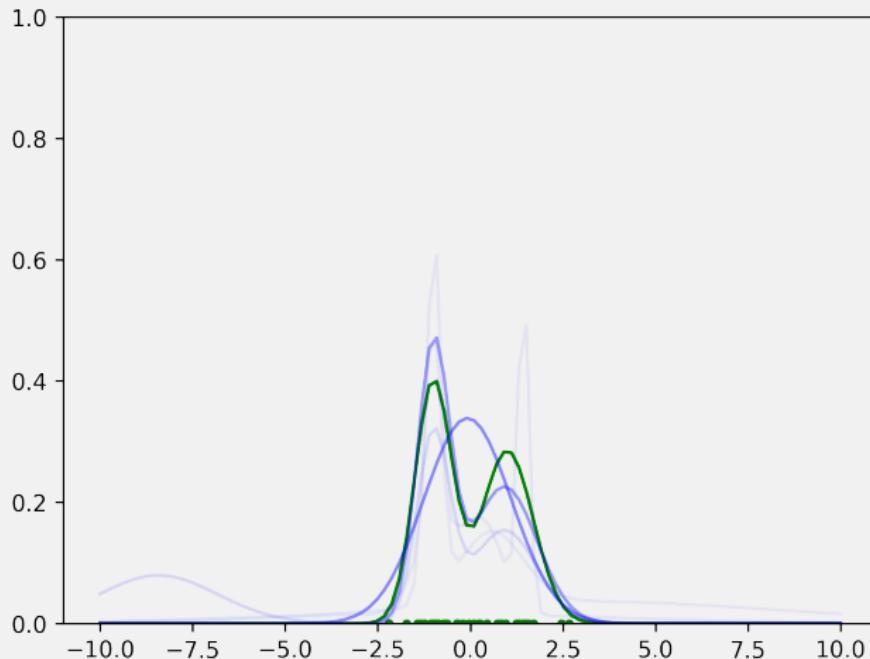
# Bayesi modell-összehasonlítás – demo

$n=43, k=1$



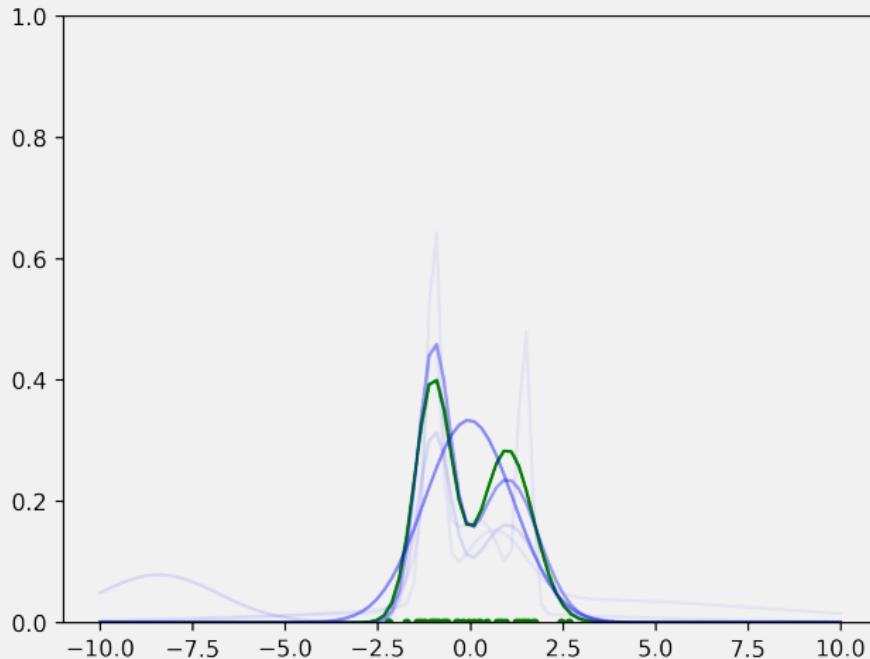
# Bayesi modell-összehasonlítás – demo

$n=44, k=1$



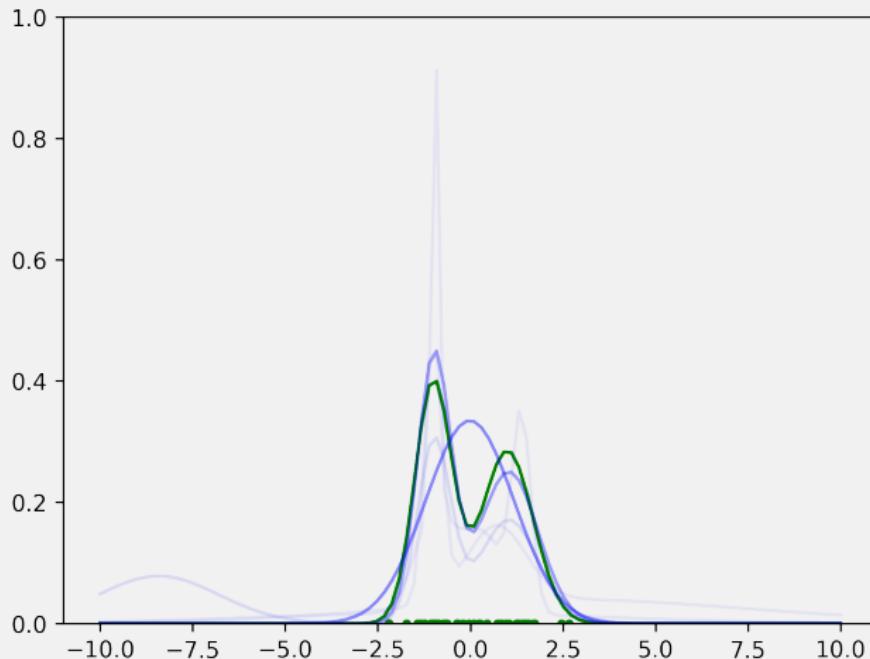
# Bayesi modell-összehasonlítás – demo

$n=45, k=1$



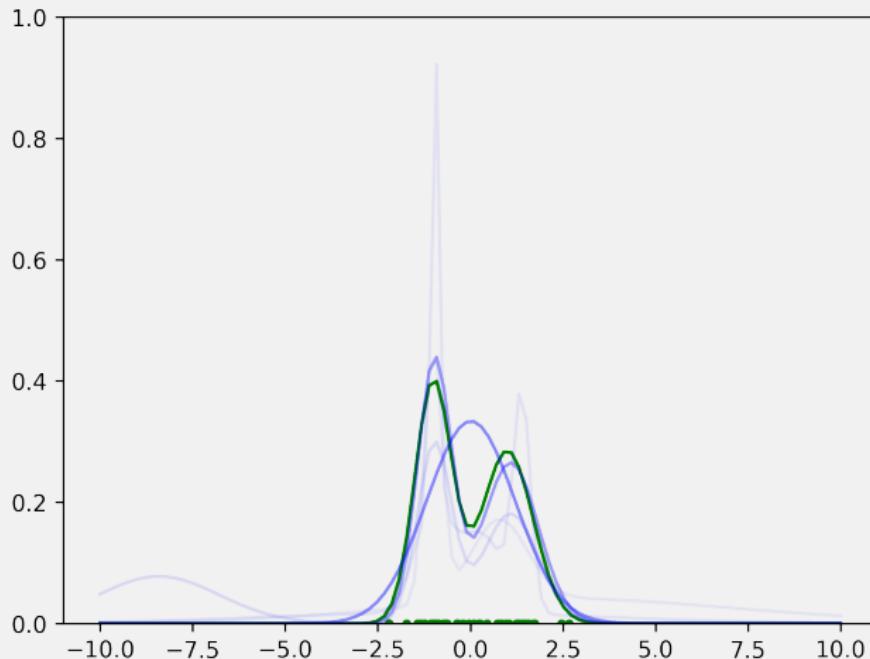
# Bayesi modell-összehasonlítás – demo

$n=46, k=1$



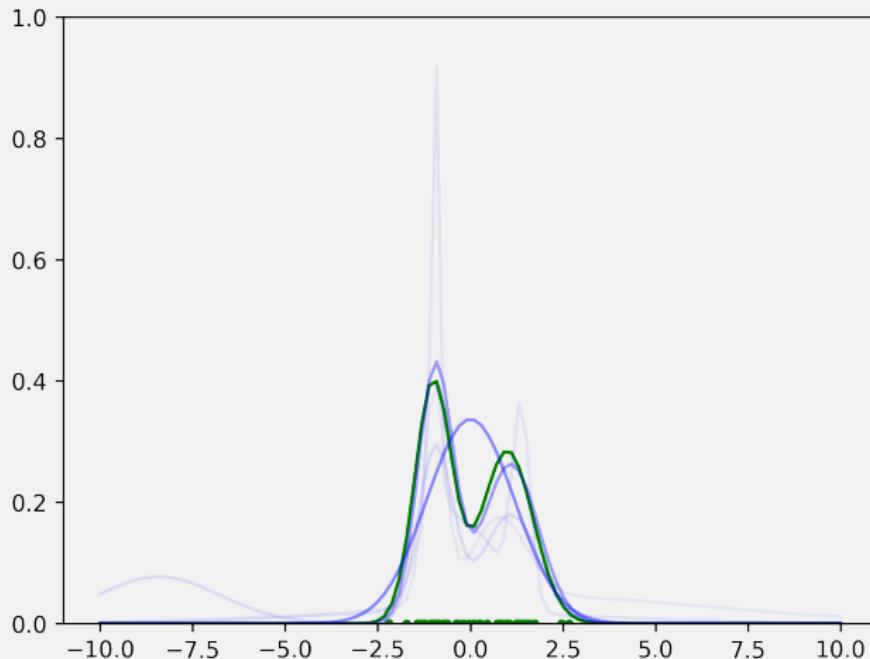
# Bayesi modell-összehasonlítás – demo

$n=47, k=1$



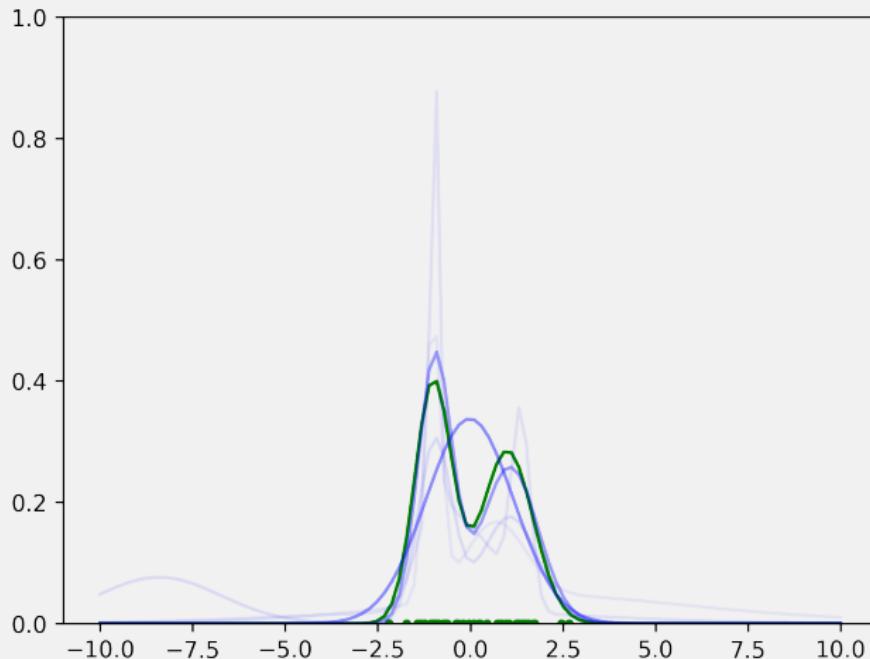
# Bayesi modell-összehasonlítás – demo

$n=48, k=1$



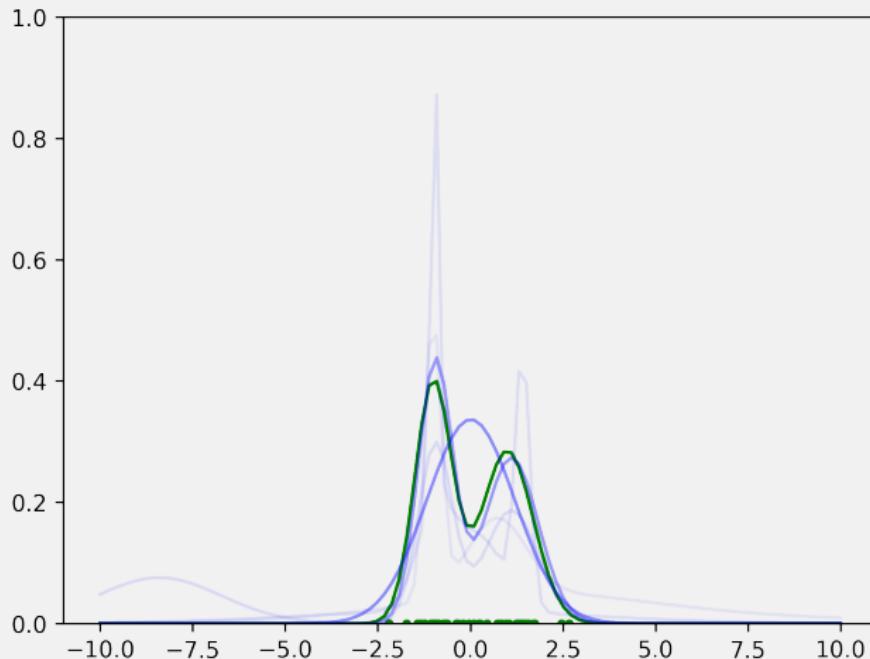
# Bayesi modell-összehasonlítás – demo

$n=49, k=1$



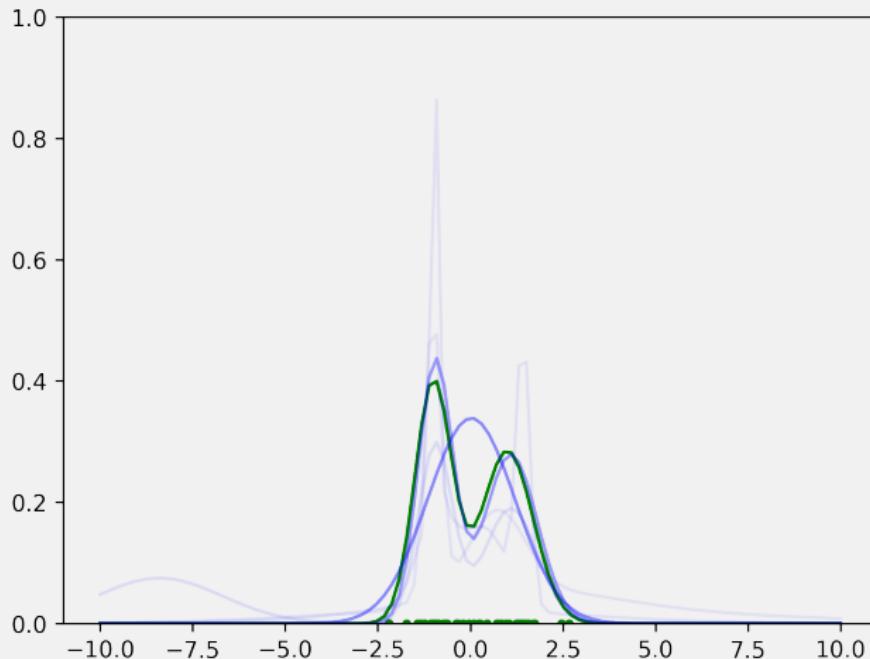
# Bayesi modell-összehasonlítás – demo

n=50, k=1



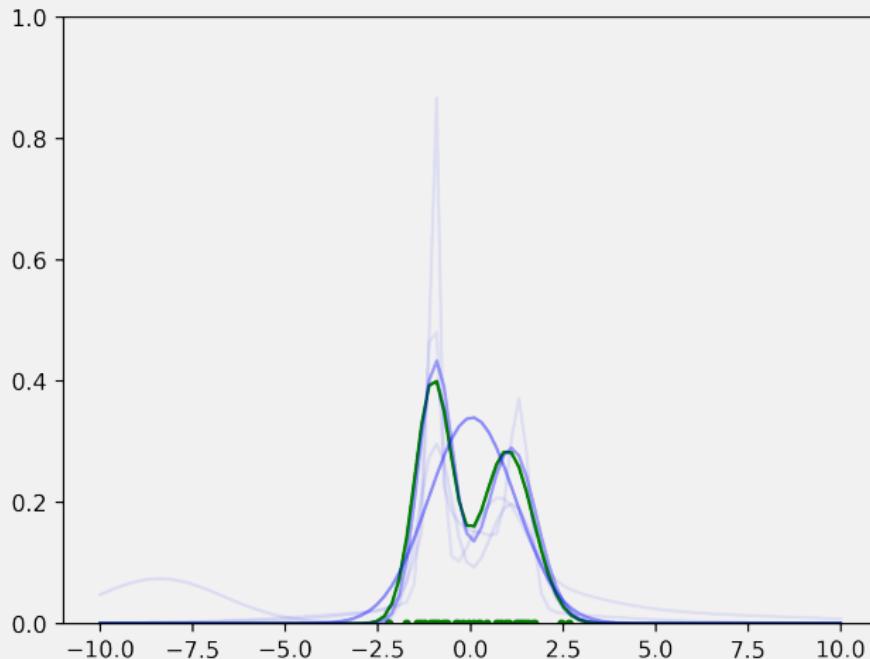
# Bayesi modell-összehasonlítás – demo

n=51, k=1



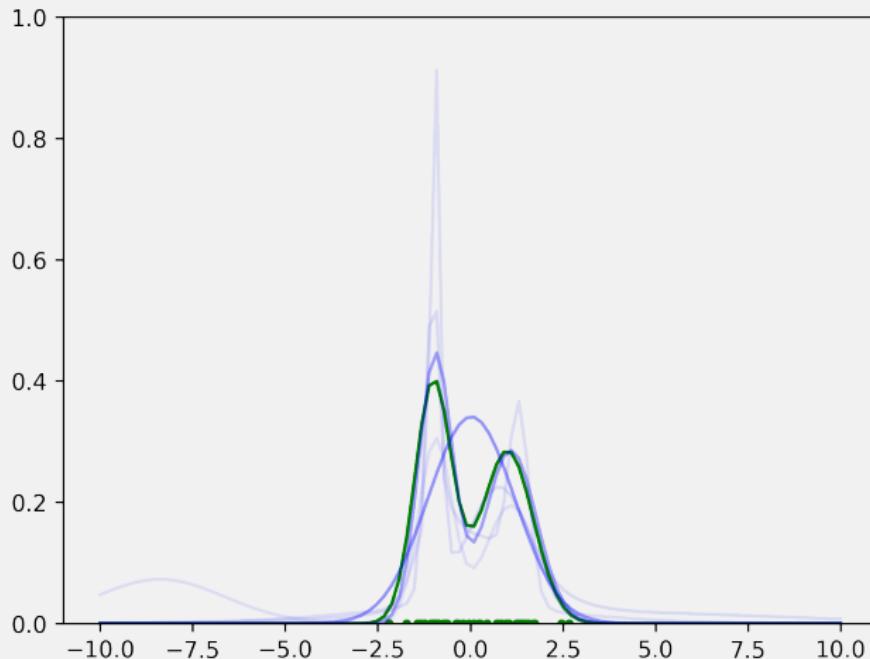
# Bayesi modell-összehasonlítás – demo

n=52, k=1



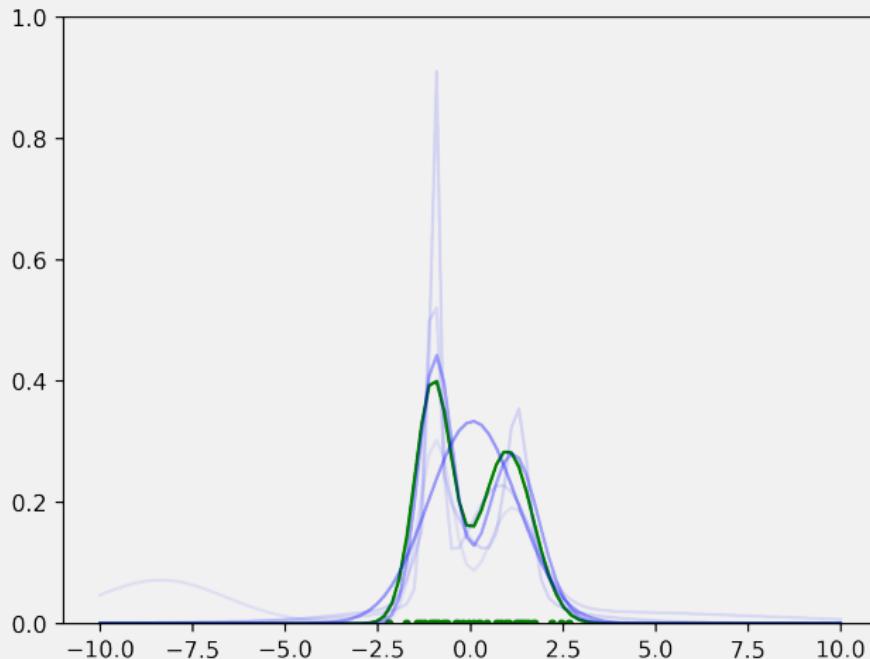
# Bayesi modell-összehasonlítás – demo

n=53, k=1



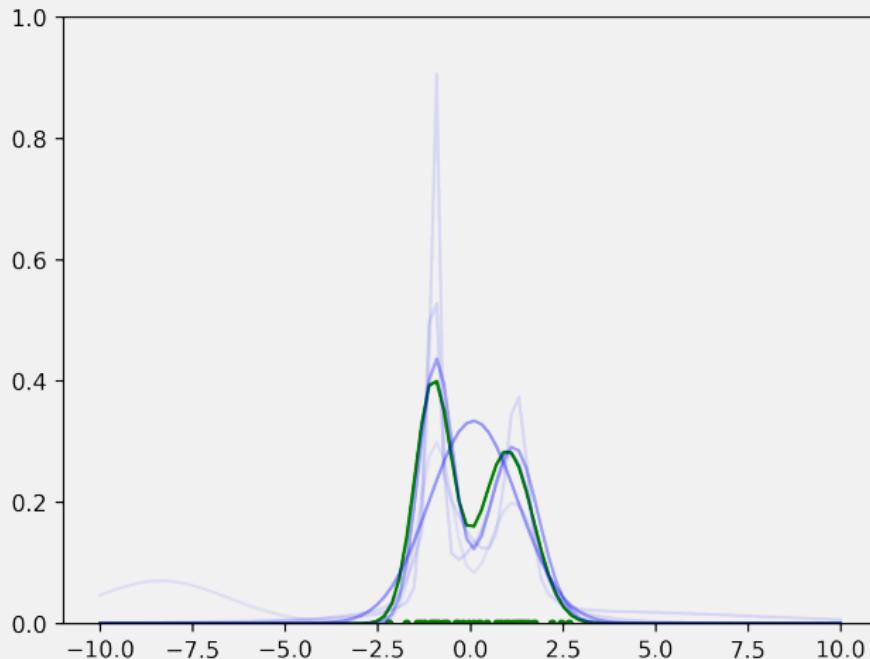
# Bayesi modell-összehasonlítás – demo

n=54, k=1



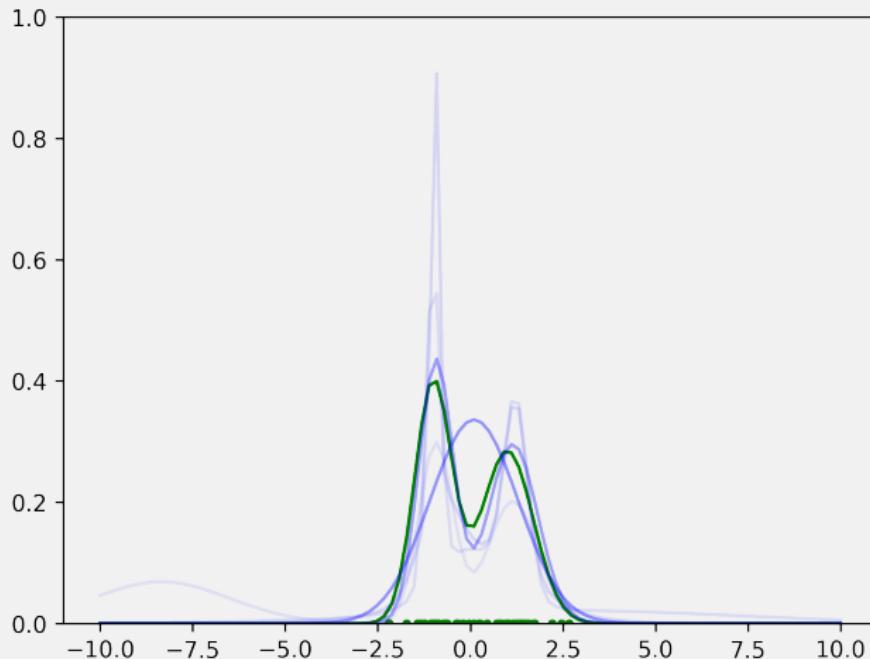
# Bayesi modell-összehasonlítás – demo

n=55, k=1



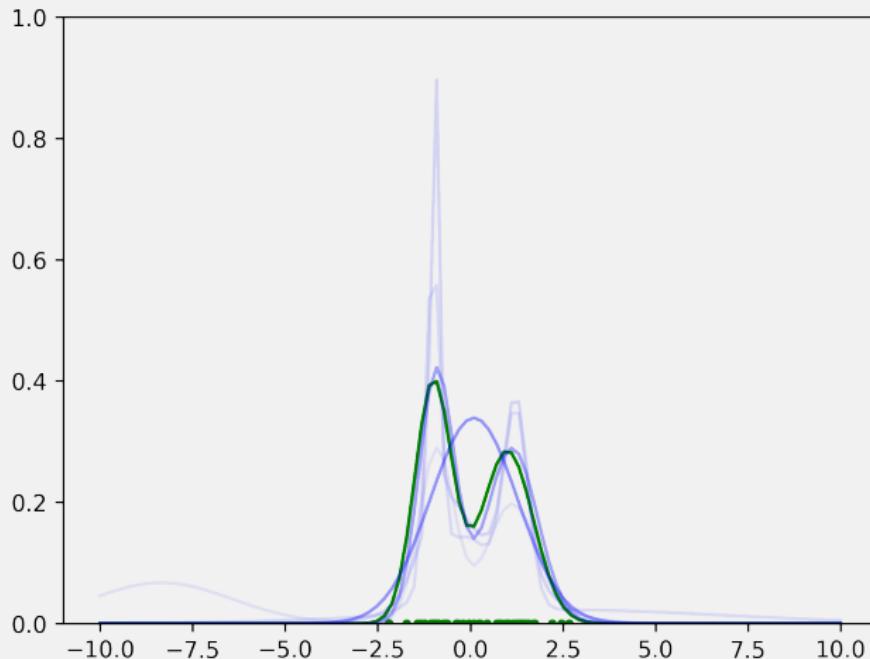
# Bayesi modell-összehasonlítás – demo

n=56, k=1



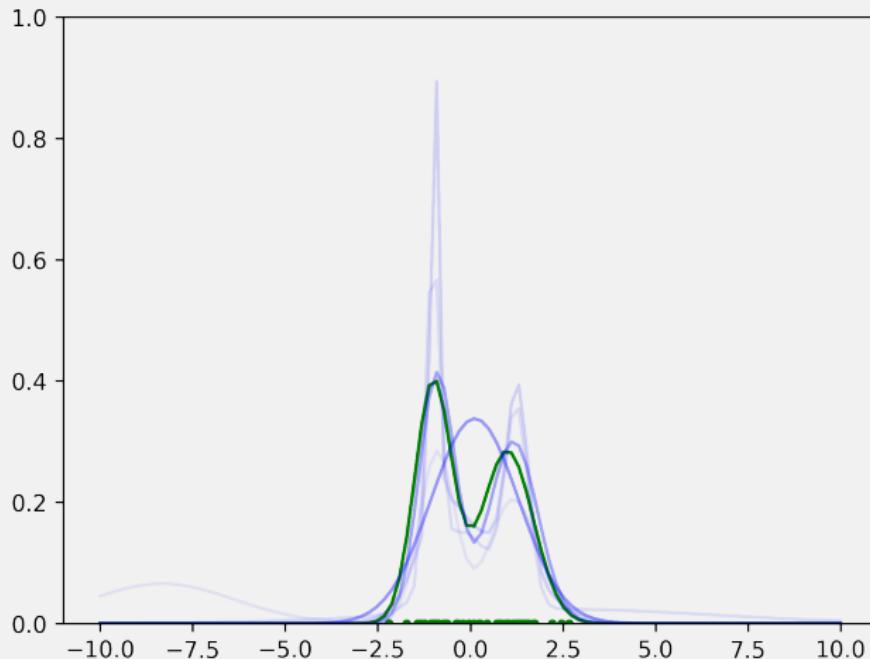
# Bayesi modell-összehasonlítás – demo

n=57, k=1



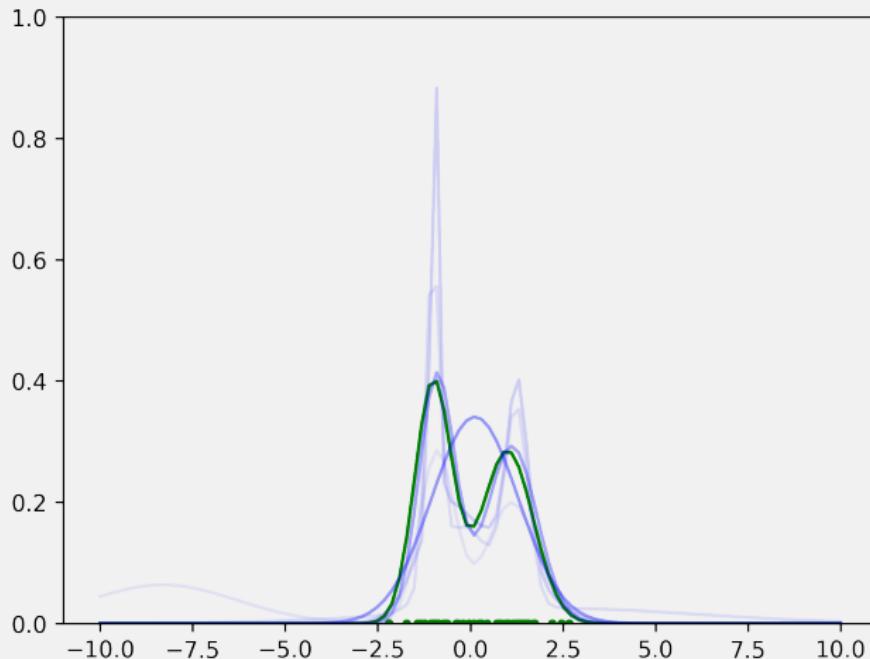
# Bayesi modell-összehasonlítás – demo

n=58, k=1



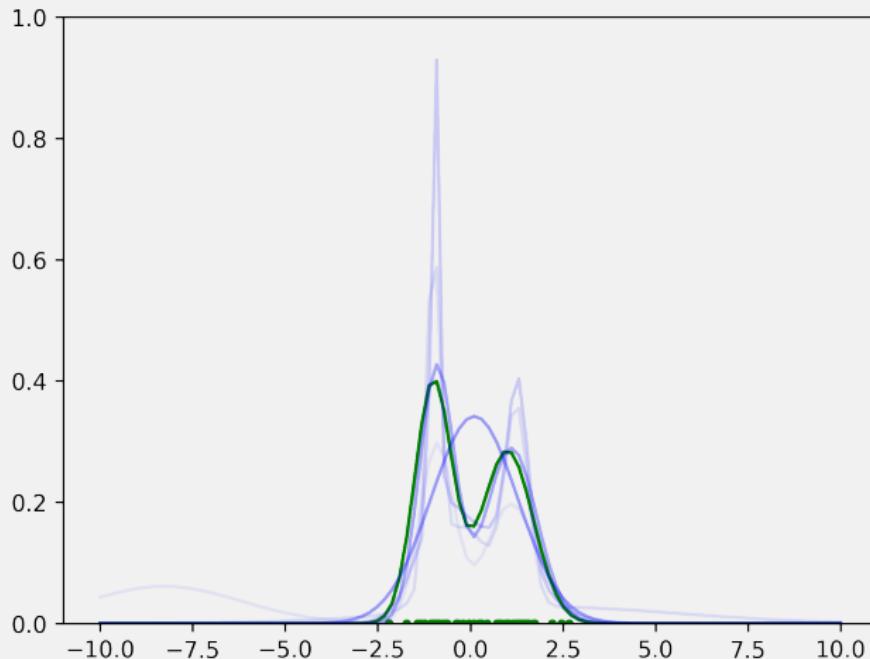
# Bayesi modell-összehasonlítás – demo

n=59, k=1



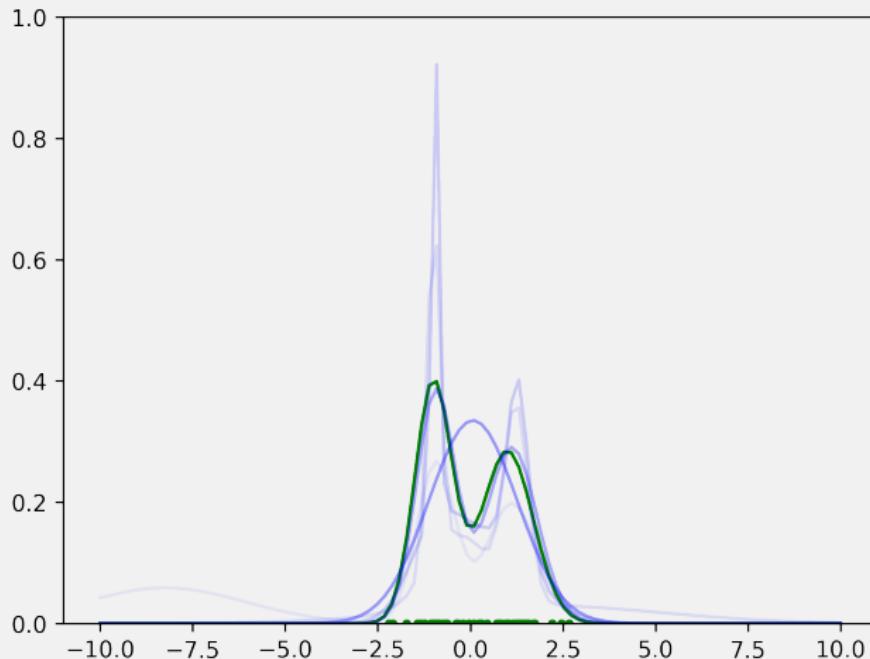
# Bayesi modell-összehasonlítás – demo

n=60, k=1



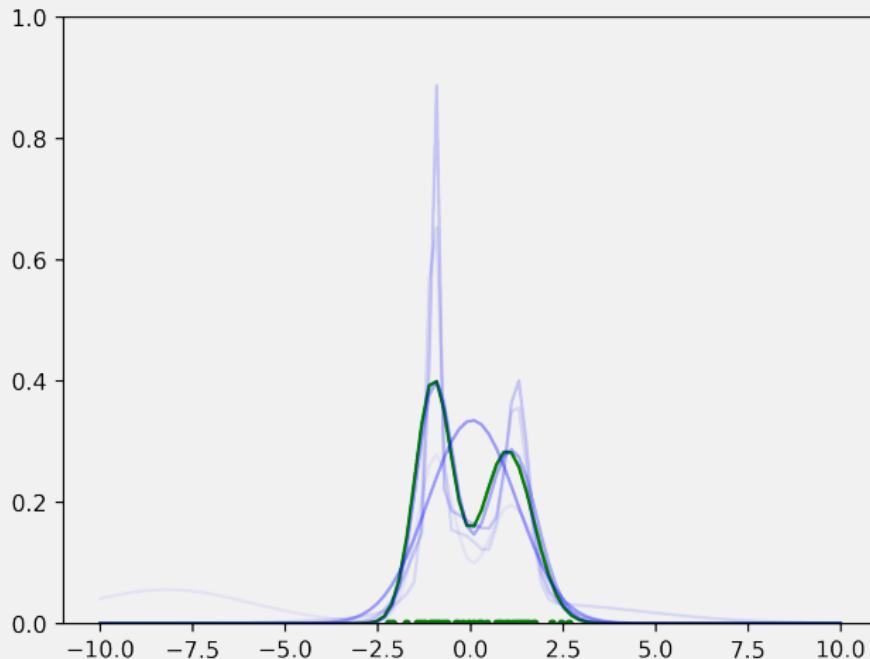
# Bayesi modell-összehasonlítás – demo

n=61, k=1



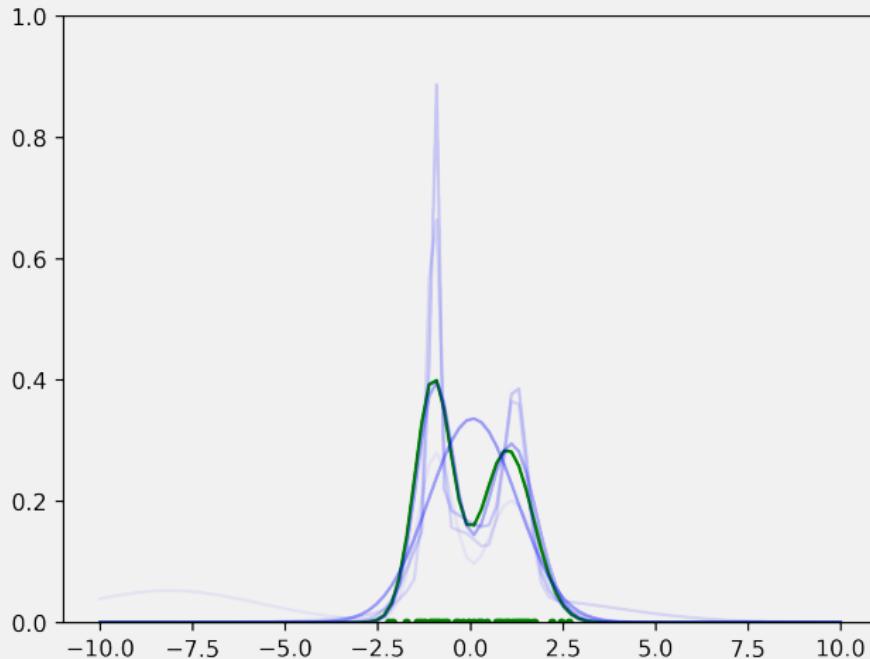
# Bayesi modell-összehasonlítás – demo

n=62, k=1



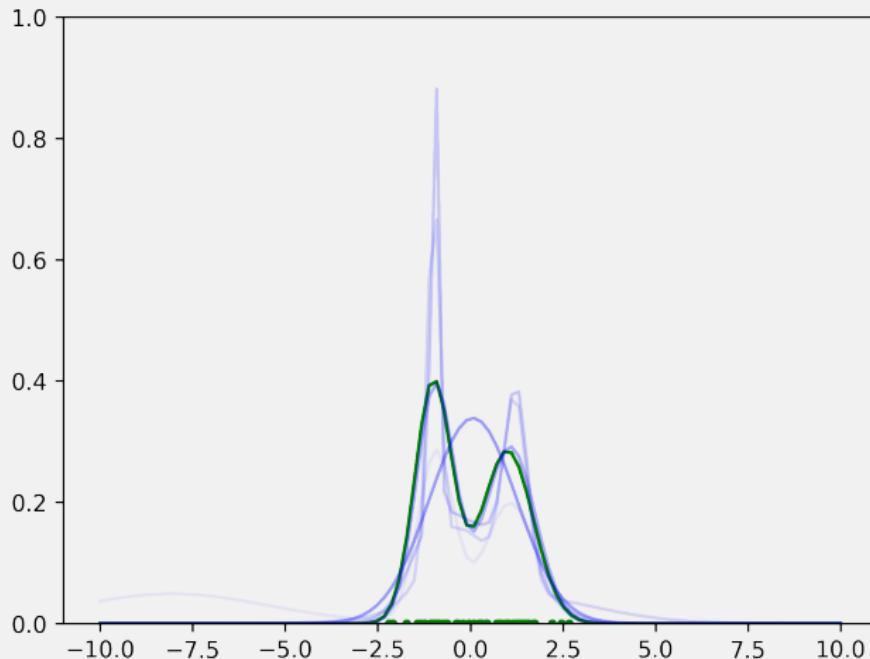
# Bayesi modell-összehasonlítás – demo

n=63, k=1



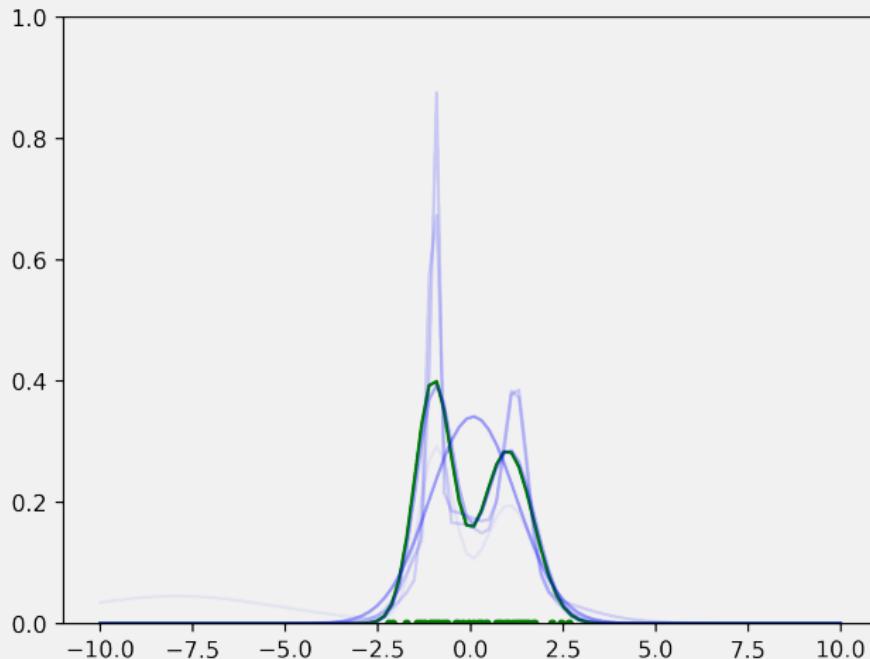
# Bayesi modell-összehasonlítás – demo

n=64, k=1



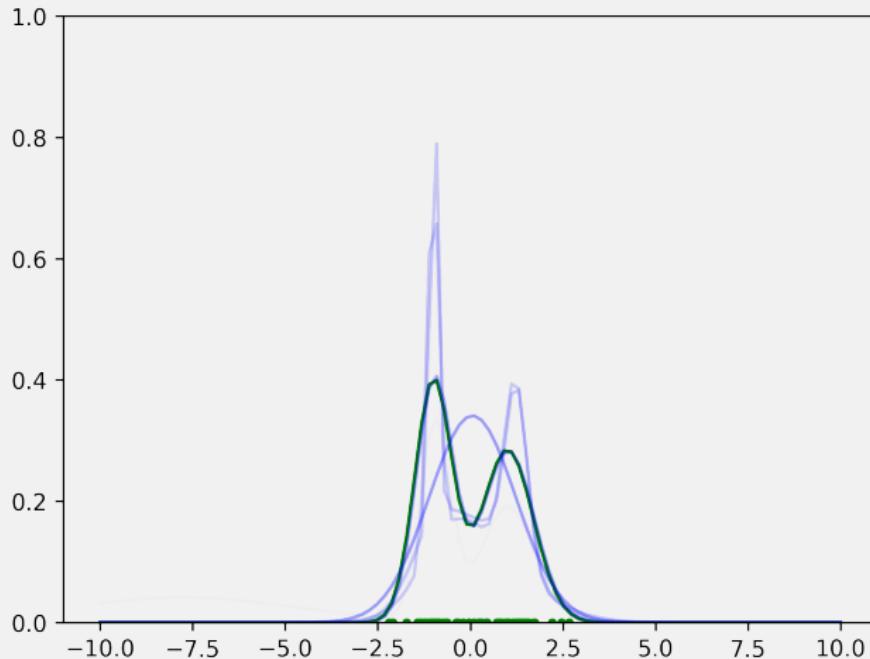
# Bayesi modell-összehasonlítás – demo

n=65, k=1



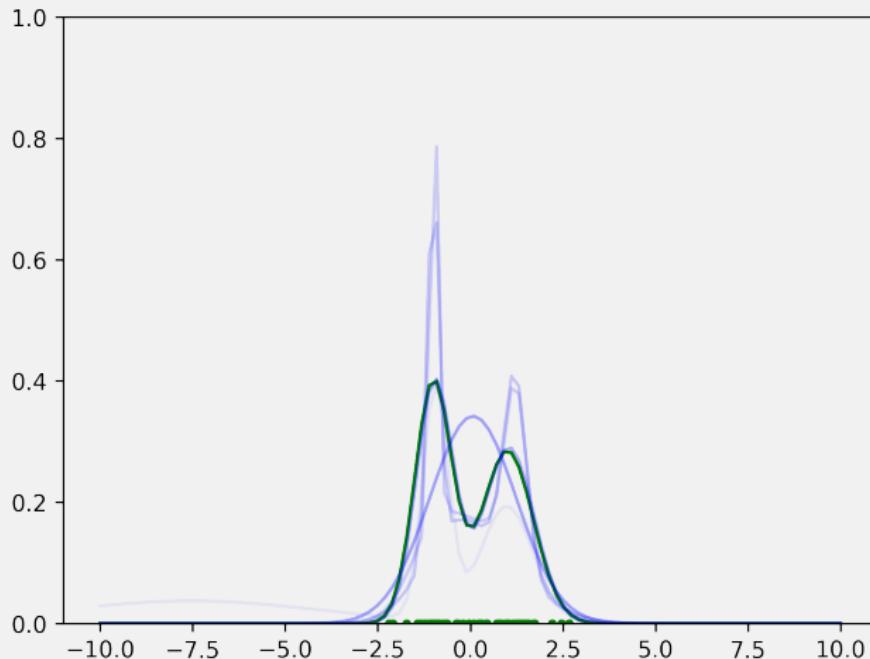
# Bayesi modell-összehasonlítás – demo

n=66, k=1



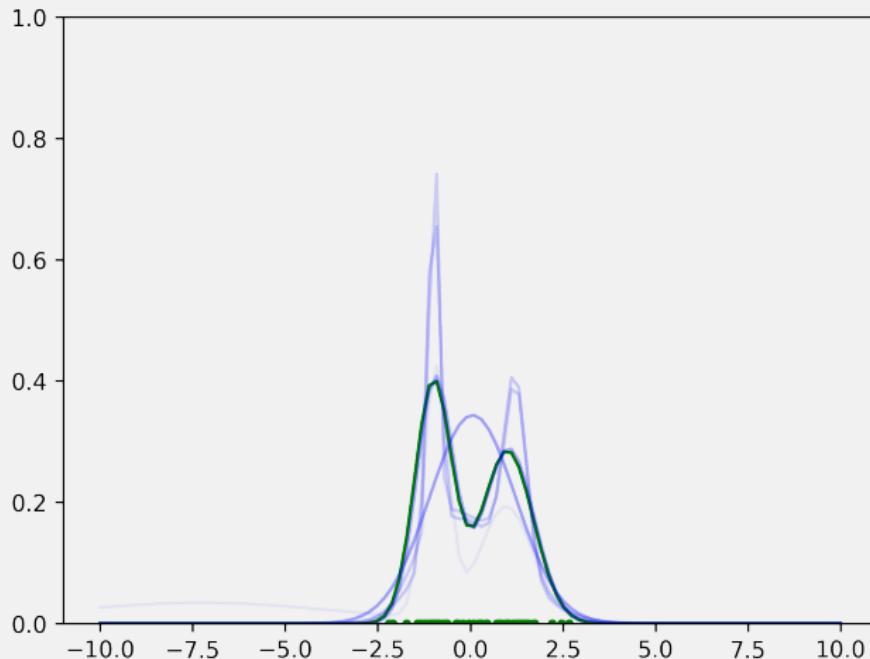
# Bayesi modell-összehasonlítás – demo

n=67, k=1



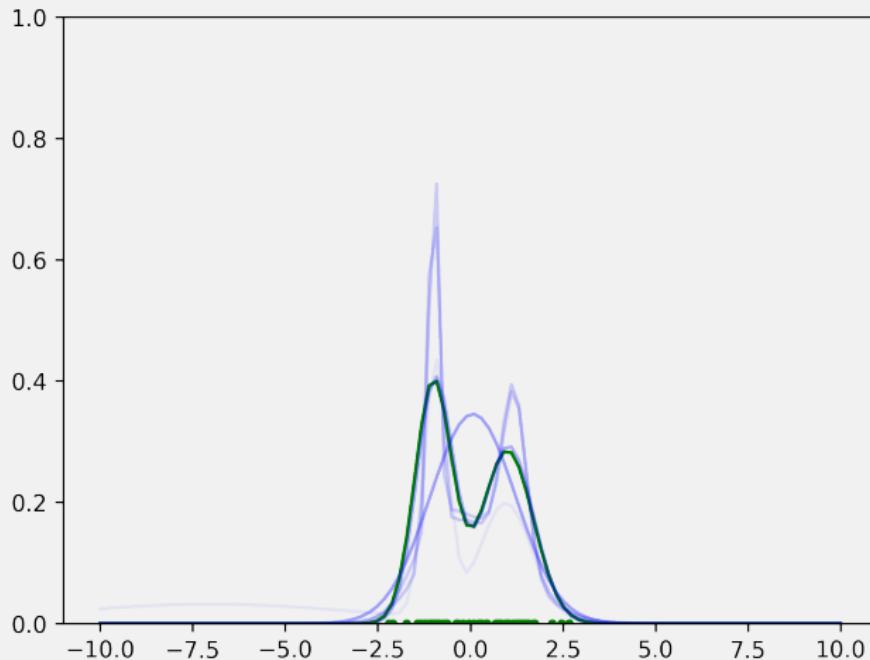
# Bayesi modell-összehasonlítás – demo

n=68, k=1



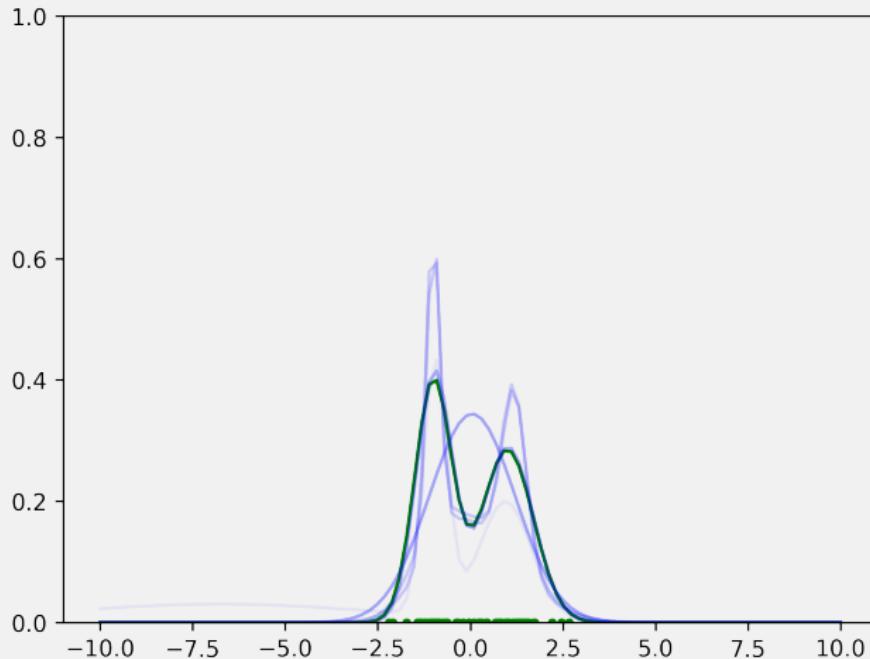
# Bayesi modell-összehasonlítás – demo

n=69, k=1



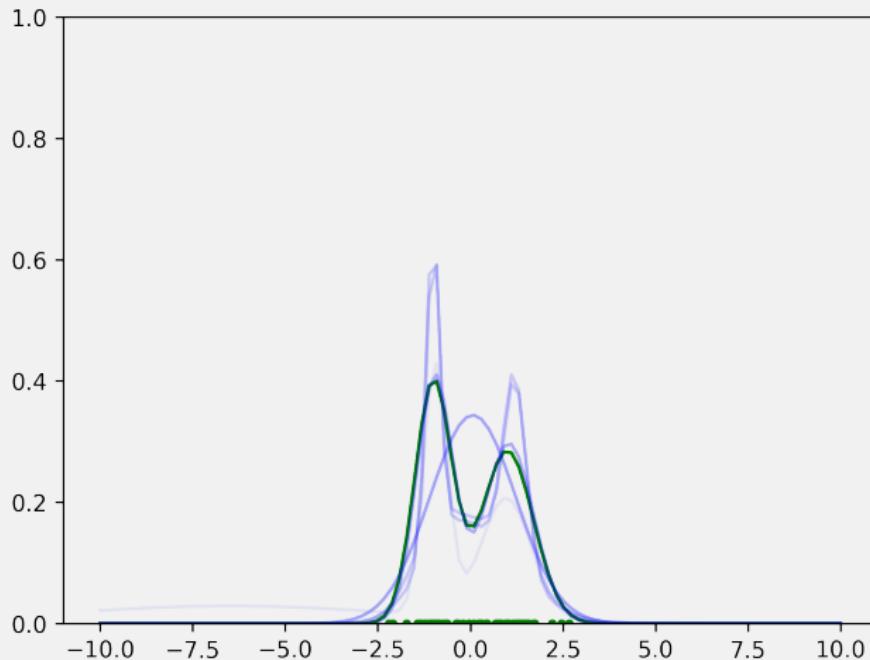
# Bayesi modell-összehasonlítás – demo

n=70, k=1



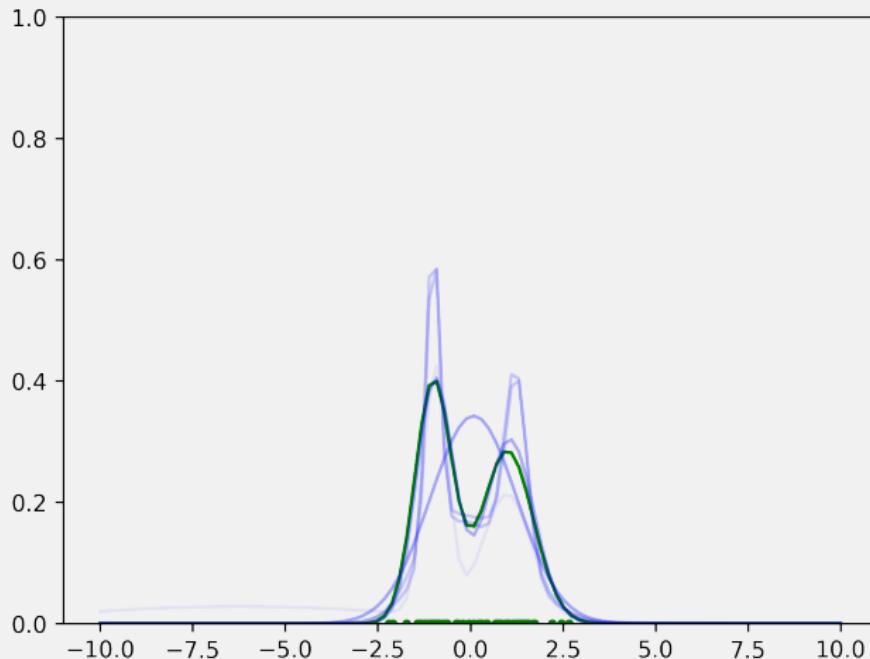
# Bayesi modell-összehasonlítás – demo

n=71, k=1



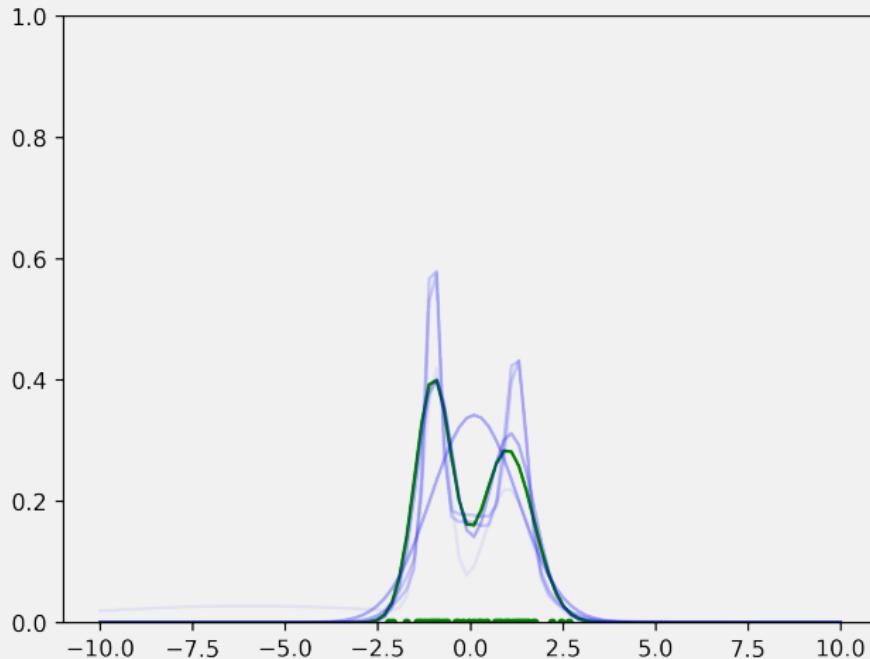
# Bayesi modell-összehasonlítás – demo

n=72, k=1



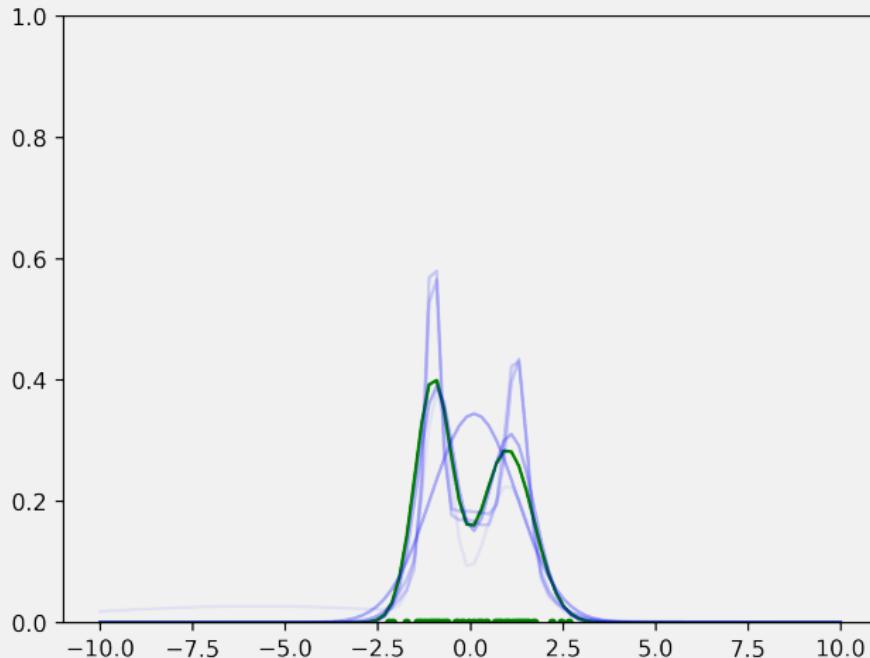
# Bayesi modell-összehasonlítás – demo

n=73, k=1



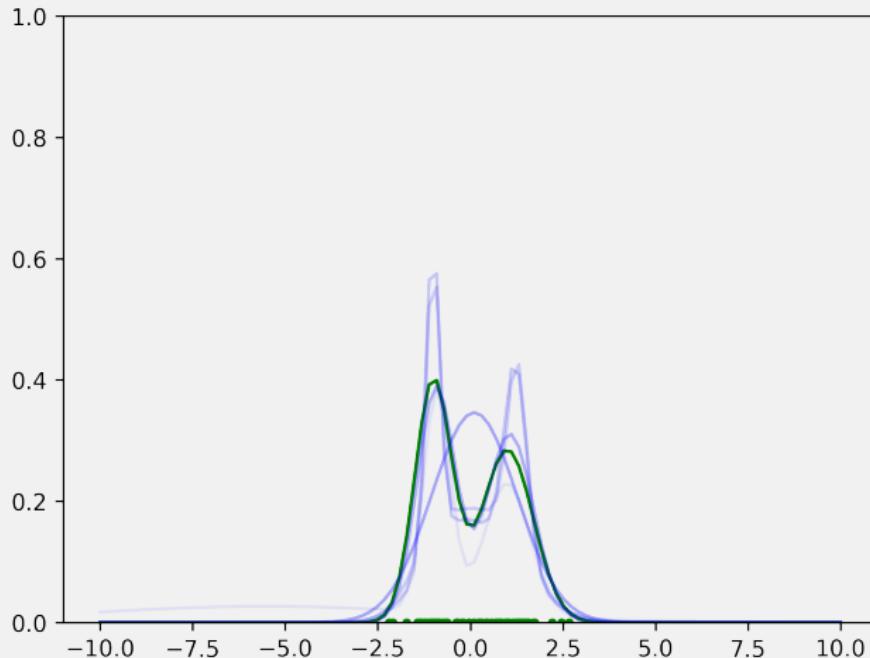
# Bayesi modell-összehasonlítás – demo

n=74, k=1



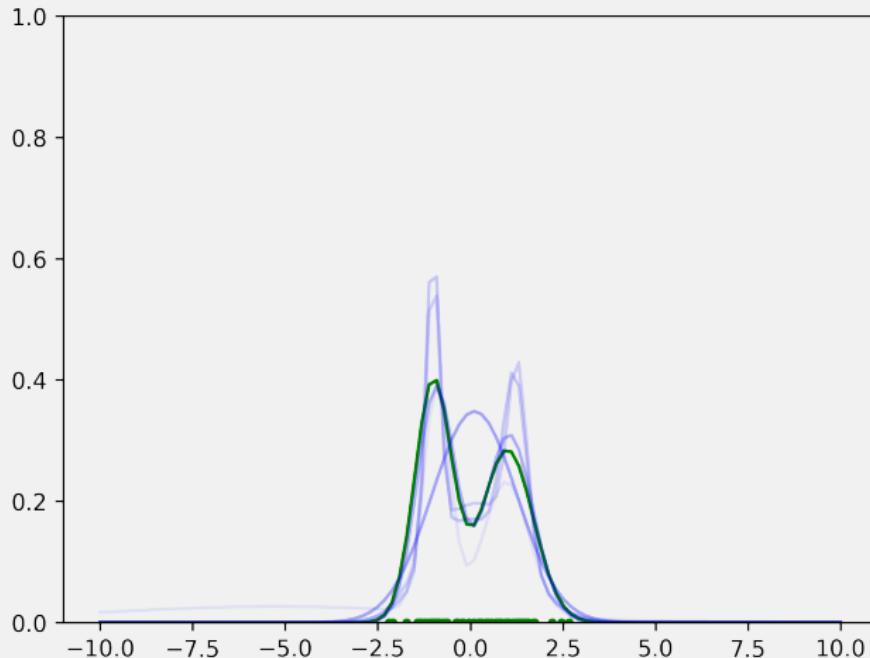
# Bayesi modell-összehasonlítás – demo

n=75, k=1



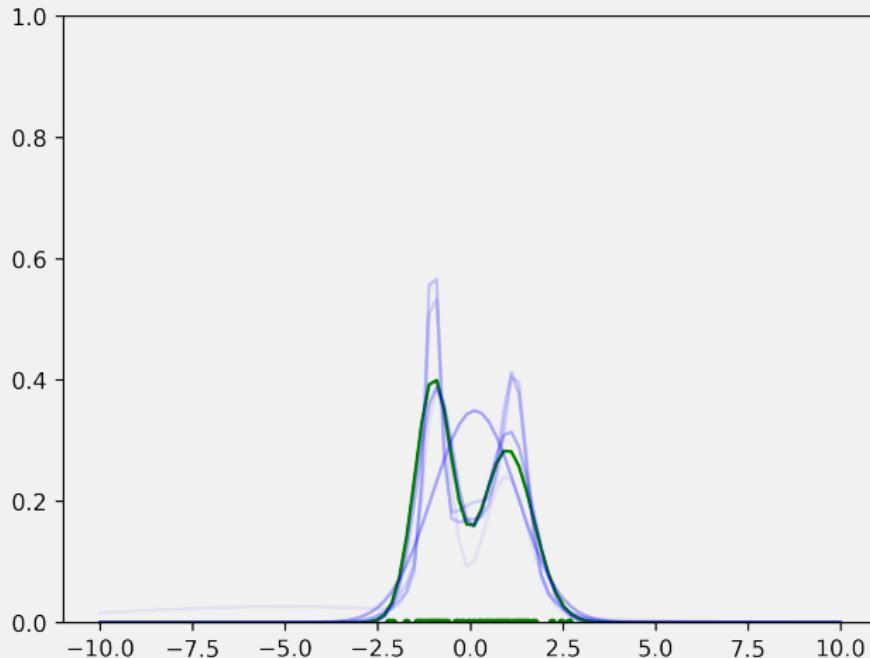
# Bayesi modell-összehasonlítás – demo

n=76, k=1



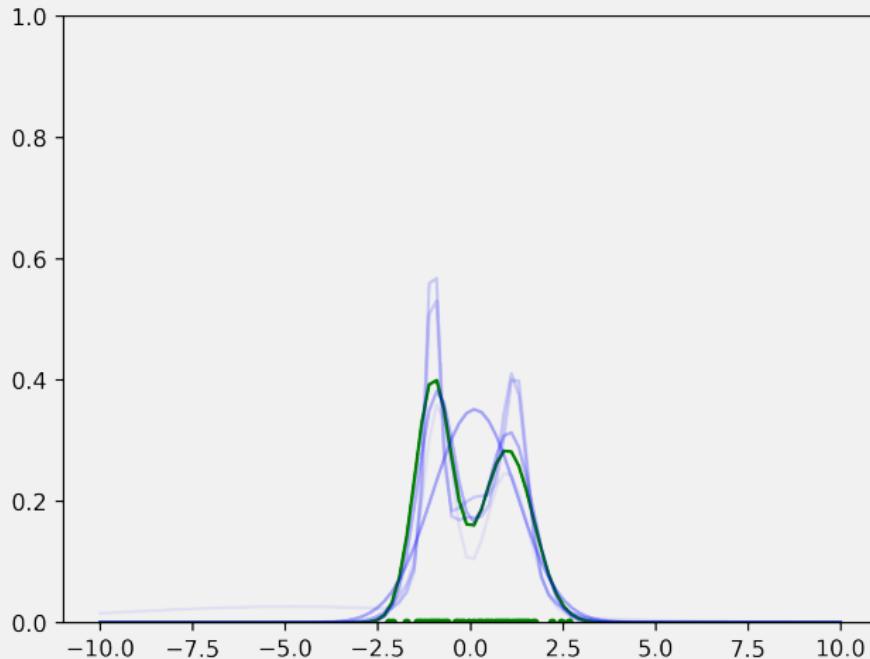
# Bayesi modell-összehasonlítás – demo

n=77, k=1



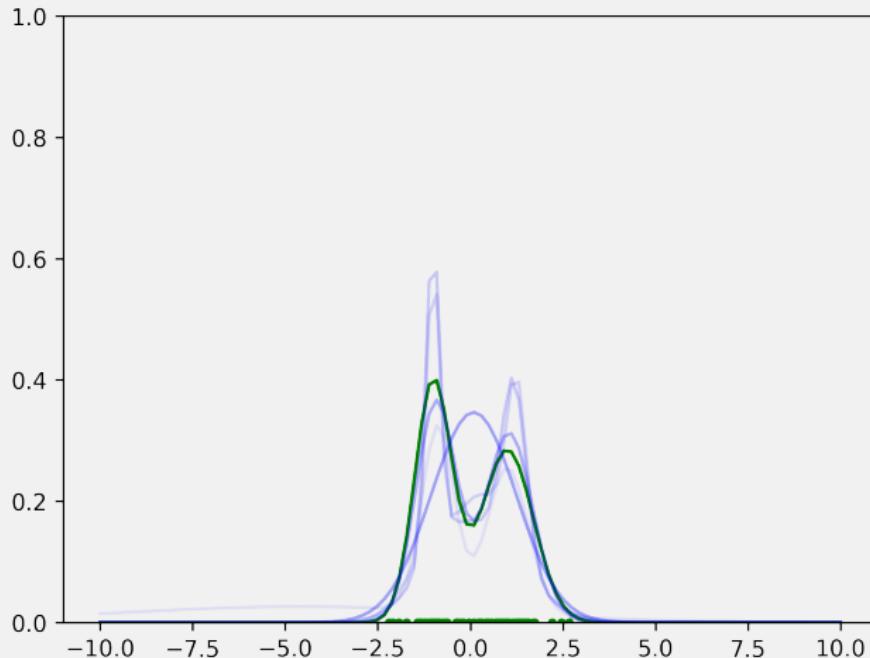
# Bayesi modell-összehasonlítás – demo

n=78, k=1



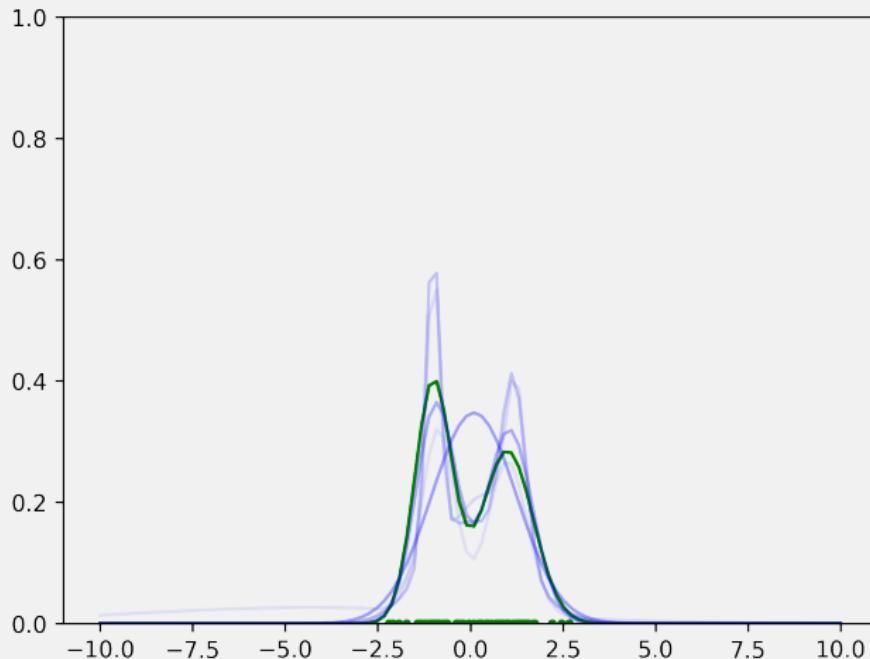
# Bayesi modell-összehasonlítás – demo

n=79, k=1



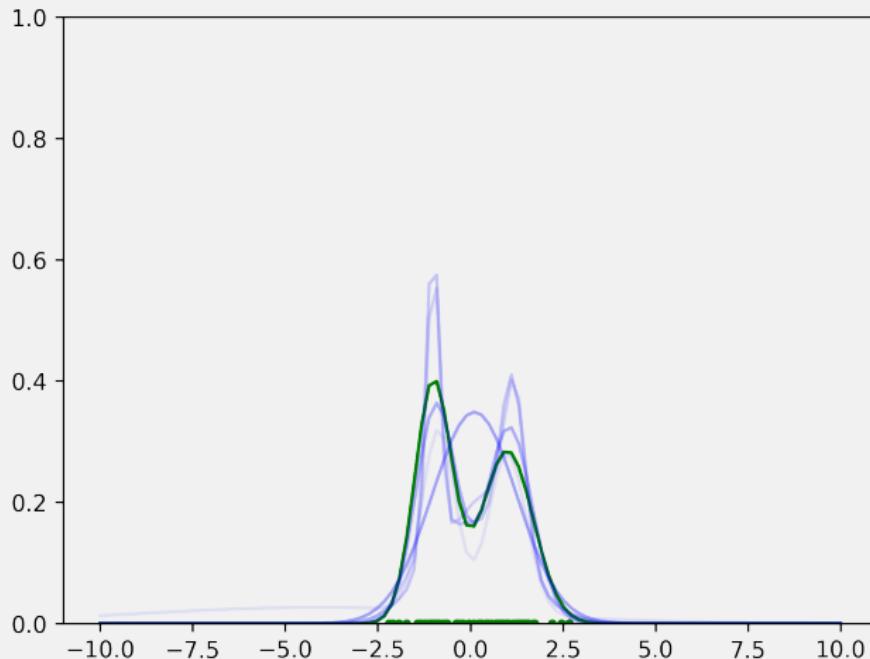
# Bayesi modell-összehasonlítás – demo

n=80, k=1



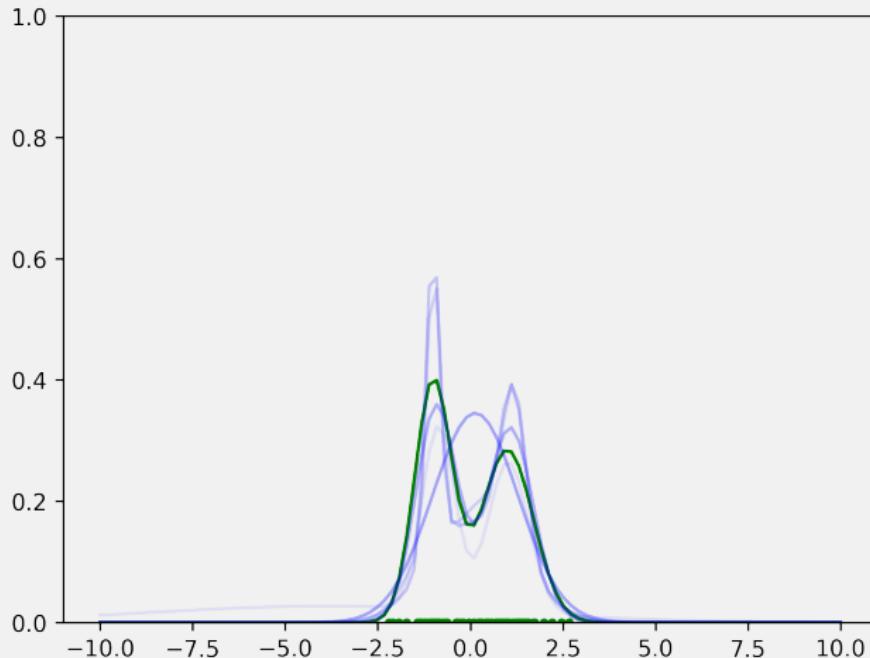
# Bayesi modell-összehasonlítás – demo

n=81, k=1



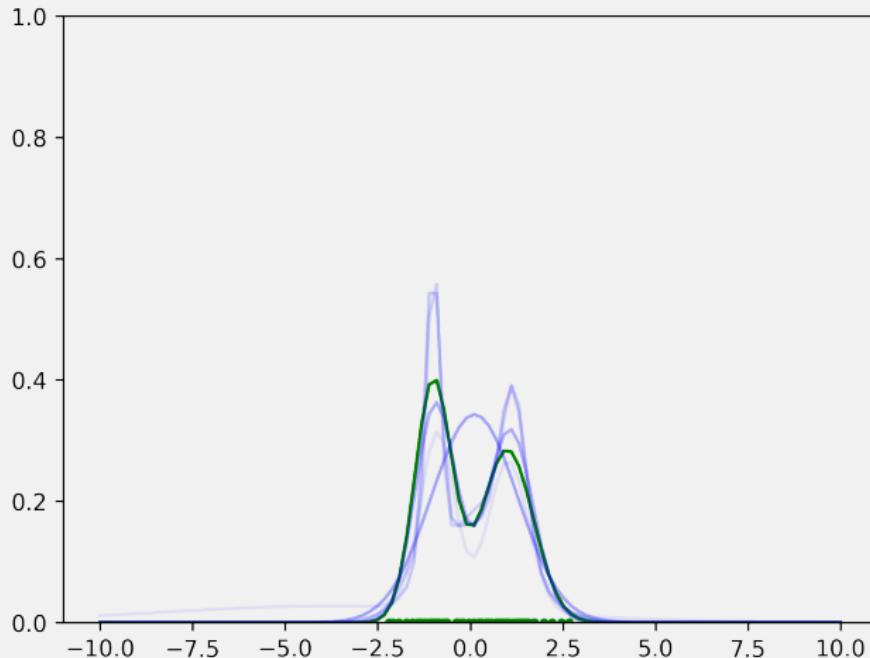
# Bayesi modell-összehasonlítás – demo

n=82, k=1



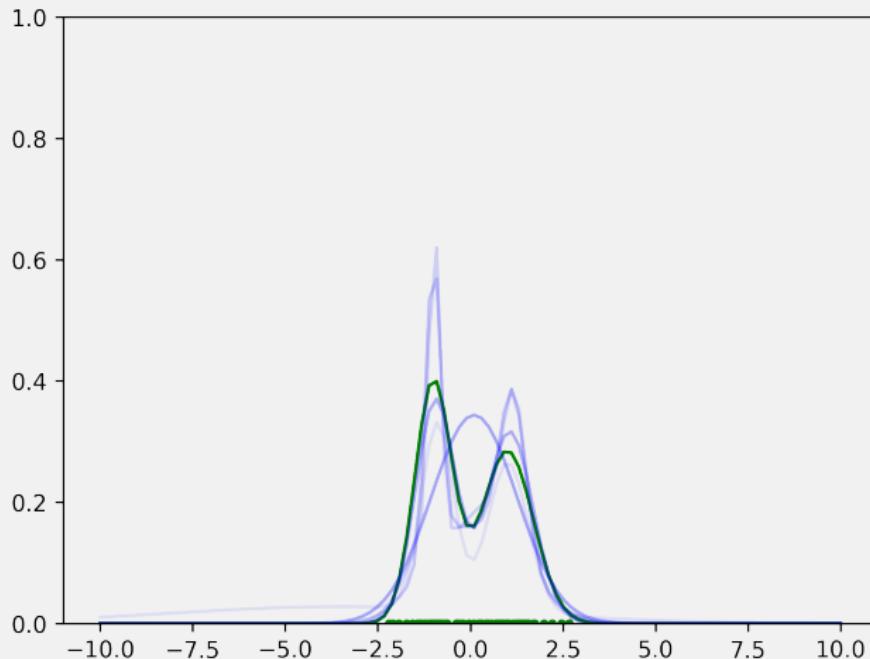
# Bayesi modell-összehasonlítás – demo

n=83, k=1



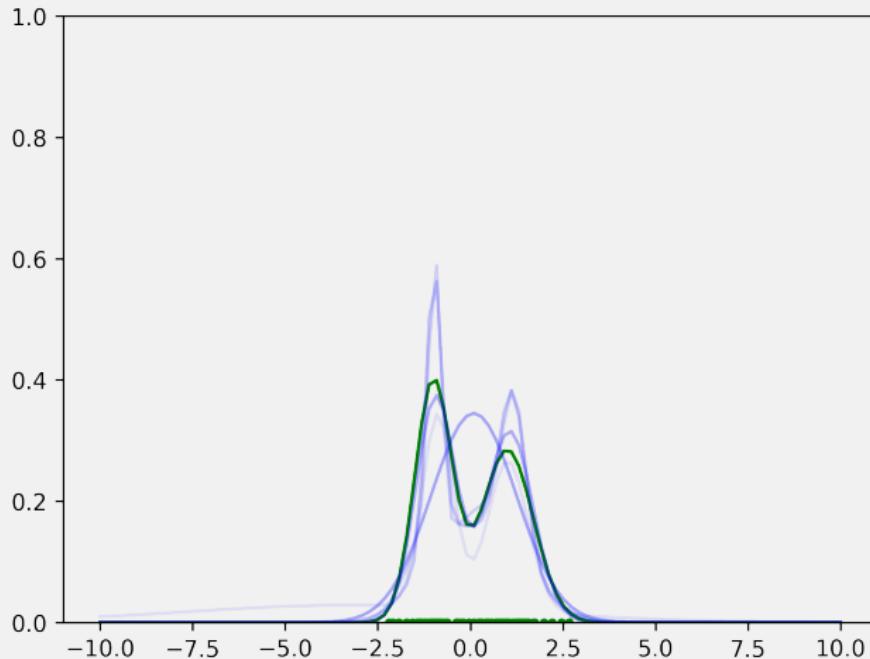
# Bayesi modell-összehasonlítás – demo

n=84, k=1



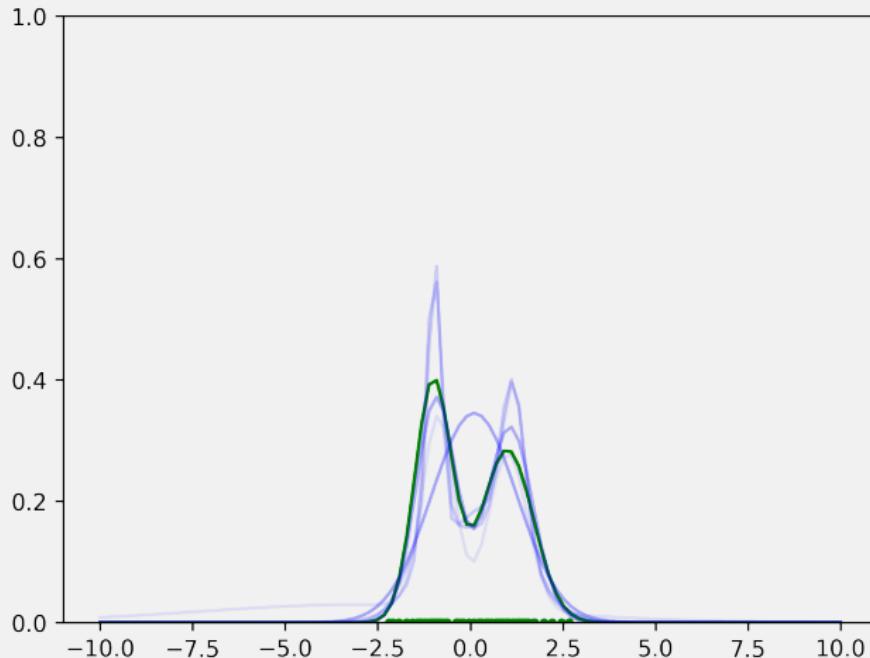
# Bayesi modell-összehasonlítás – demo

n=85, k=1



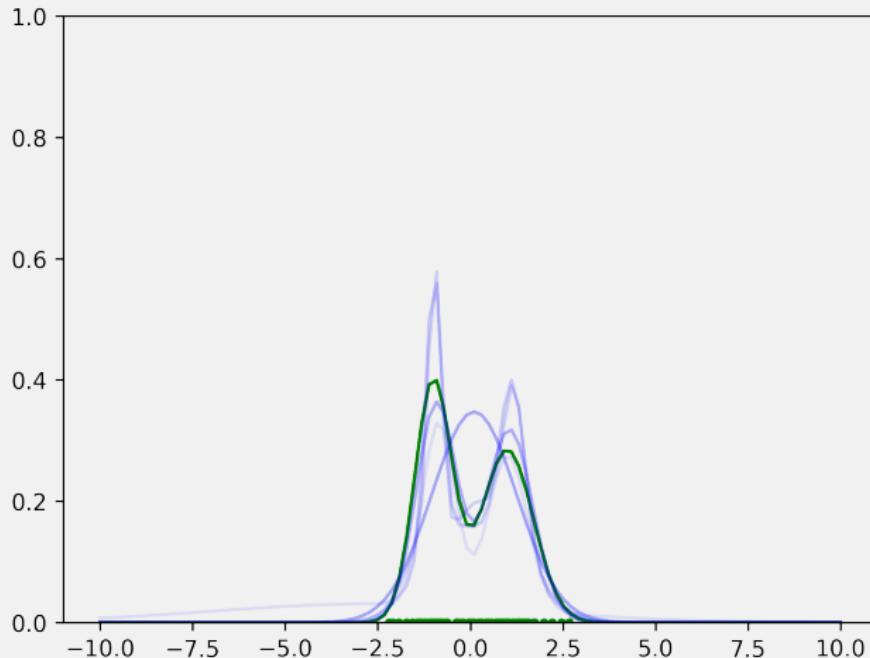
# Bayesi modell-összehasonlítás – demo

n=86, k=1



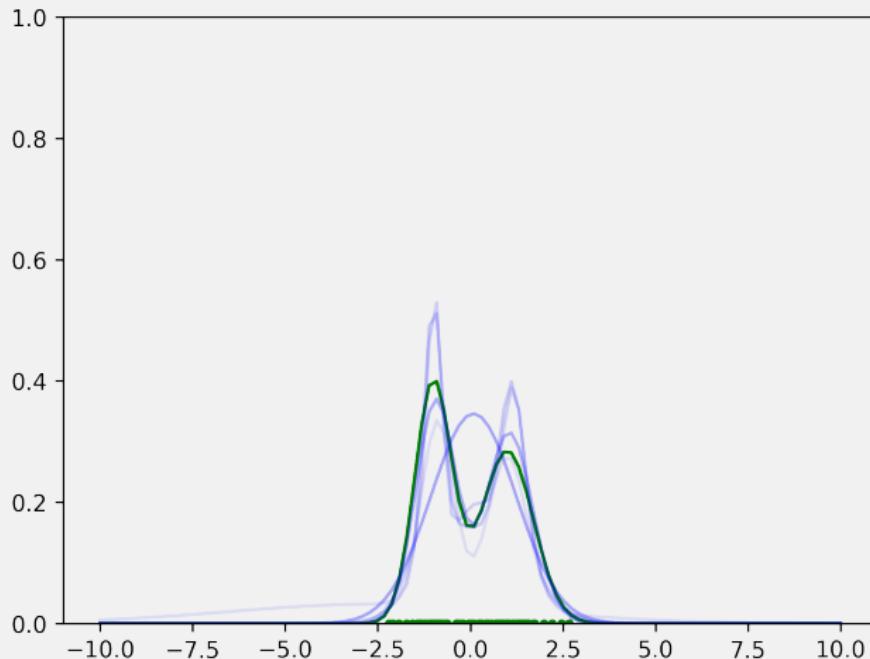
# Bayesi modell-összehasonlítás – demo

n=87, k=1



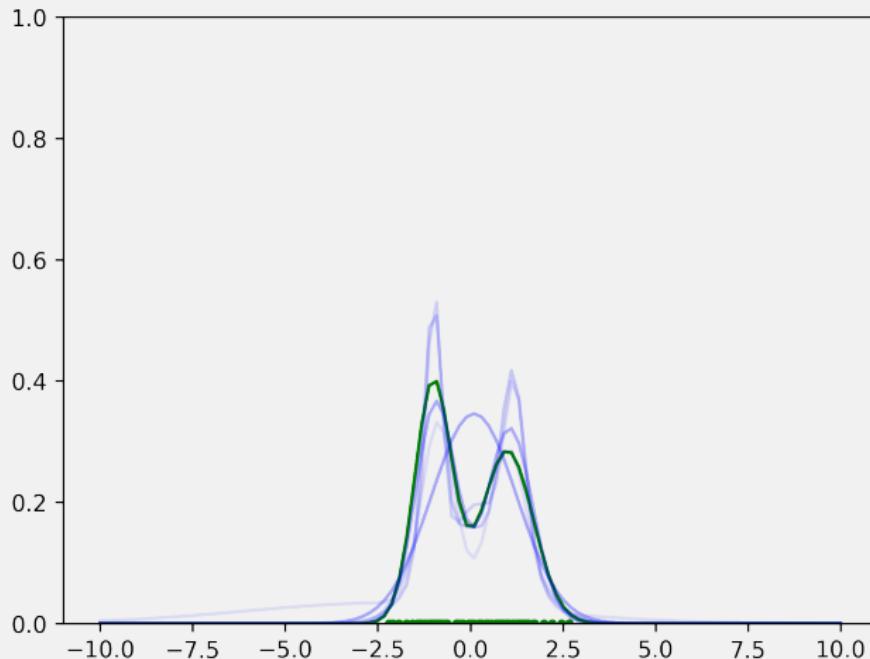
# Bayesi modell-összehasonlítás – demo

n=88, k=1



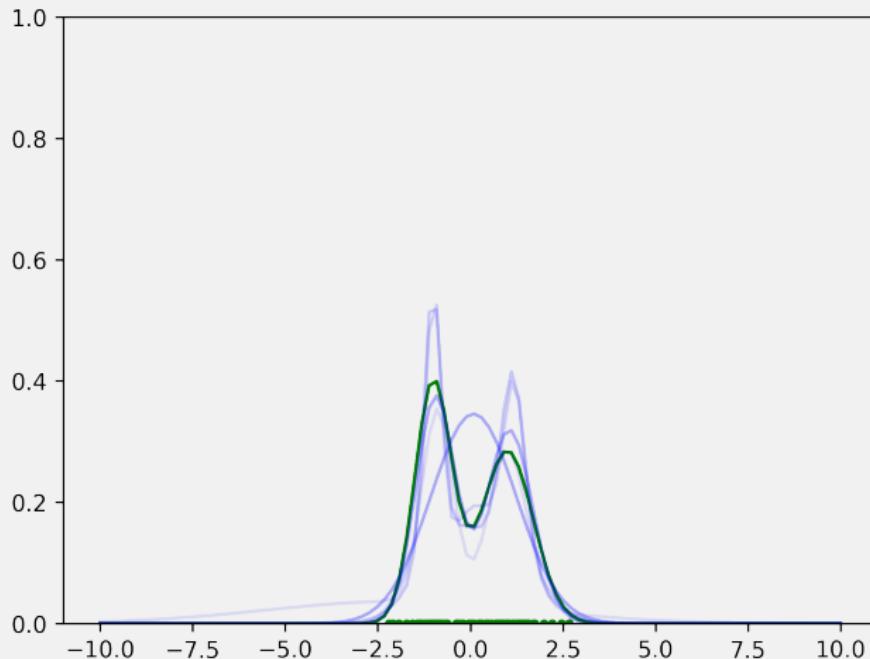
# Bayesi modell-összehasonlítás – demo

n=89, k=1



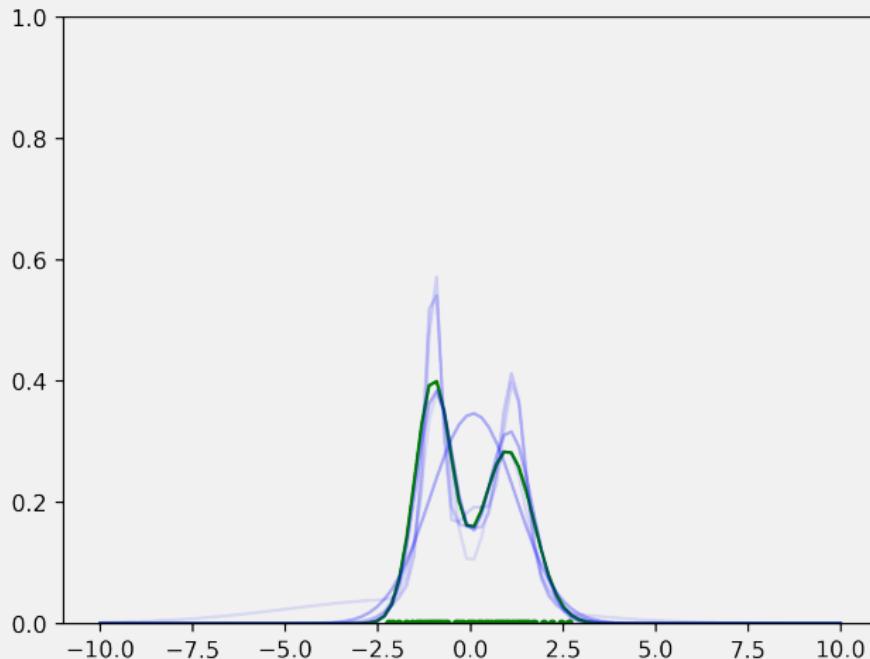
# Bayesi modell-összehasonlítás – demo

n=90, k=1



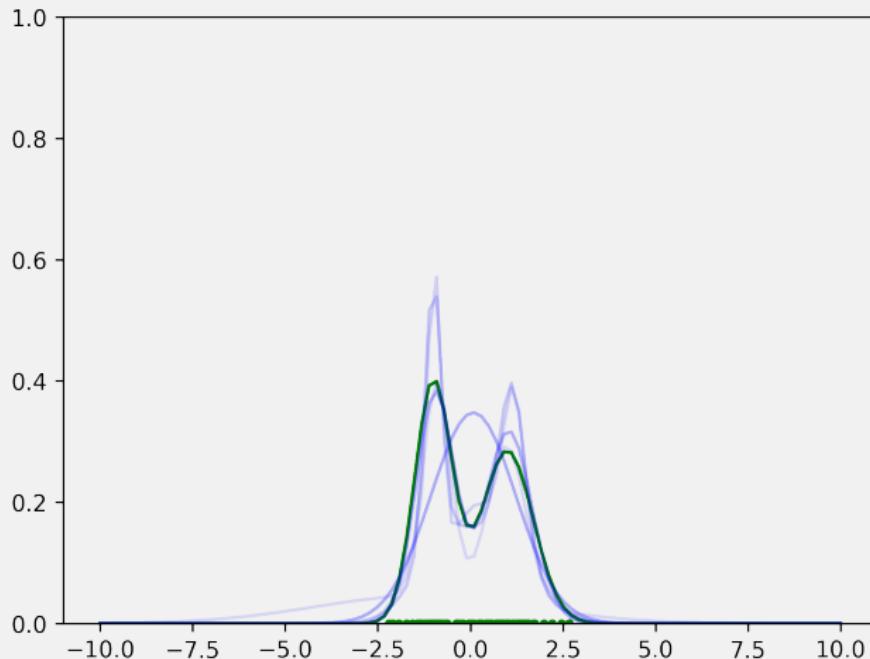
# Bayesi modell-összehasonlítás – demo

n=91, k=1



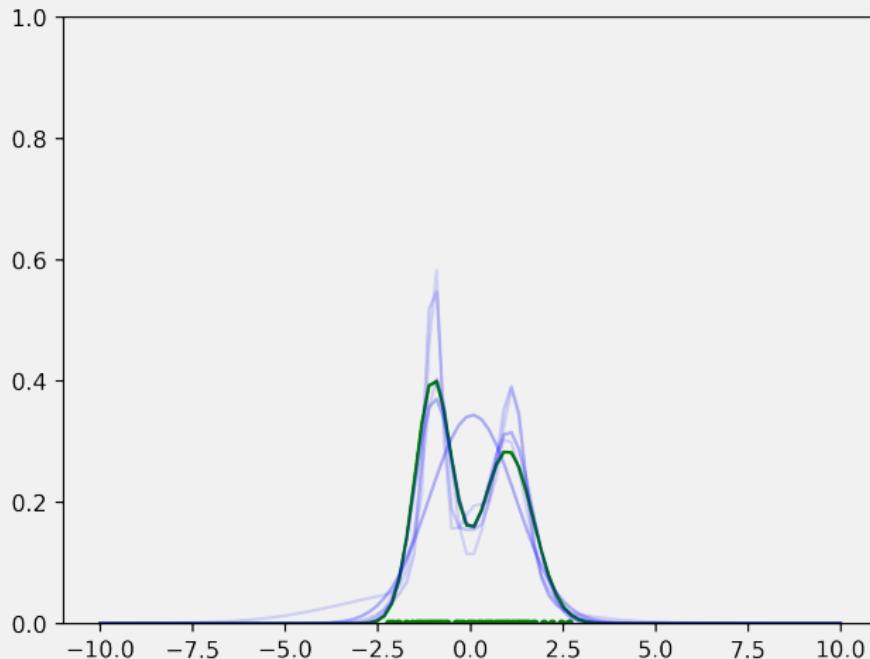
# Bayesi modell-összehasonlítás – demo

n=92, k=1



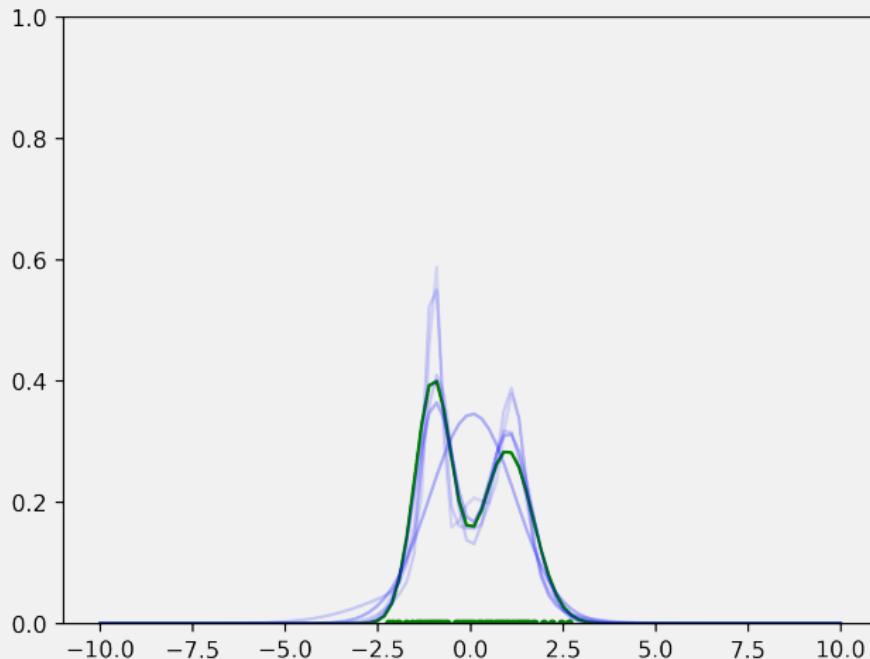
# Bayesi modell-összehasonlítás – demo

n=93, k=1



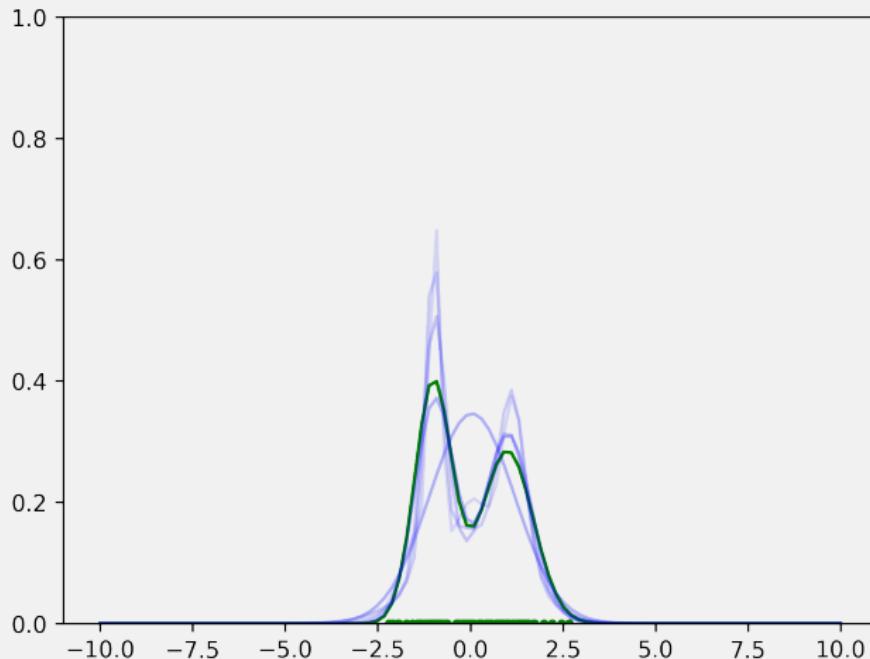
# Bayesi modell-összehasonlítás – demo

n=94, k=1



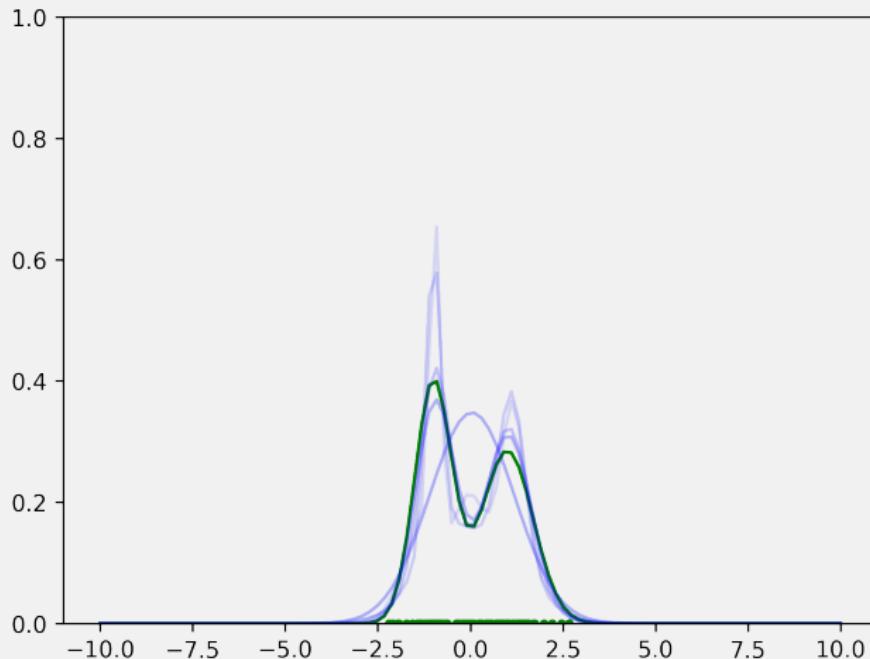
# Bayesi modell-összehasonlítás – demo

n=95, k=1



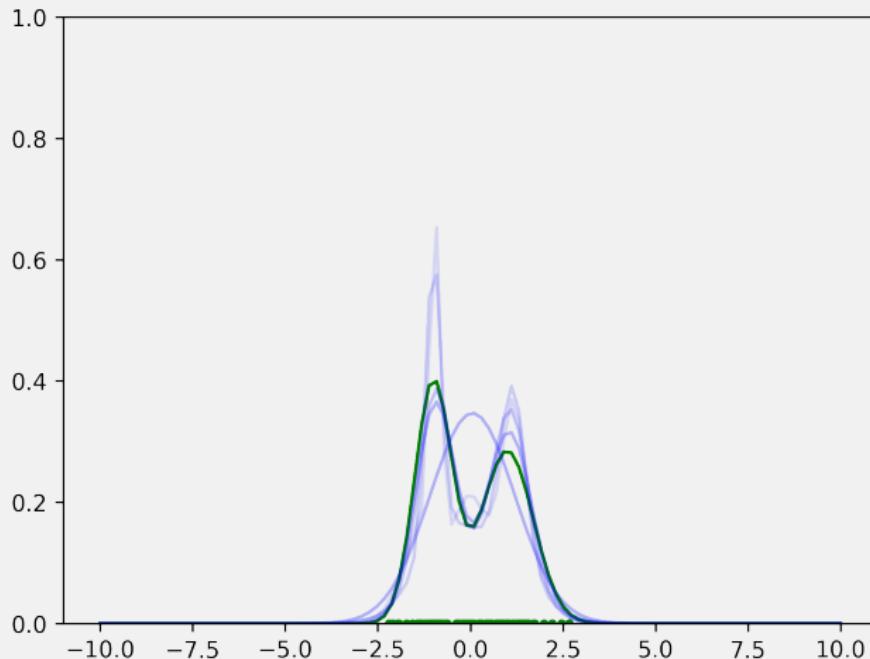
# Bayesi modell-összehasonlítás – demo

n=96, k=1



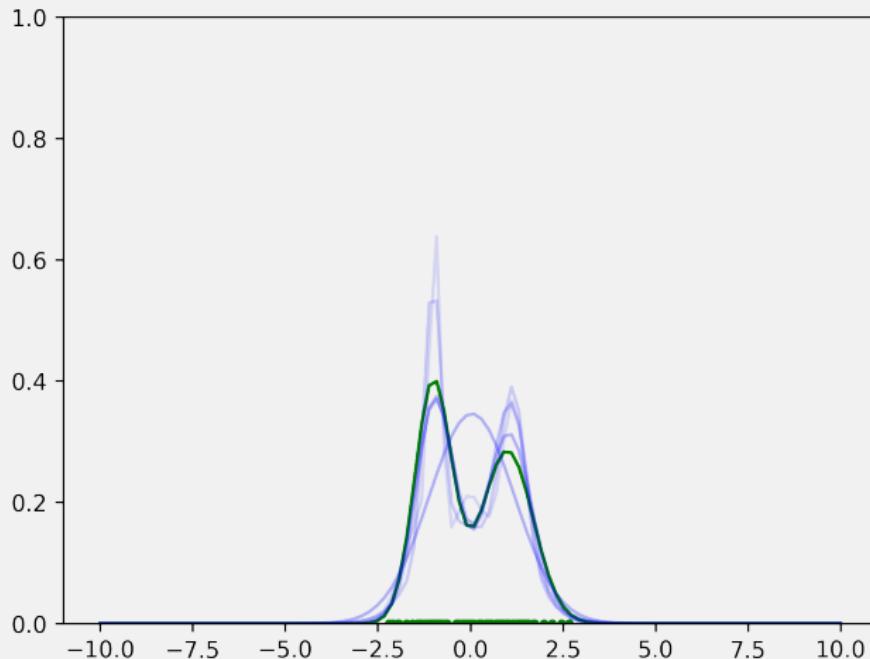
# Bayesi modell-összehasonlítás – demo

n=97, k=1



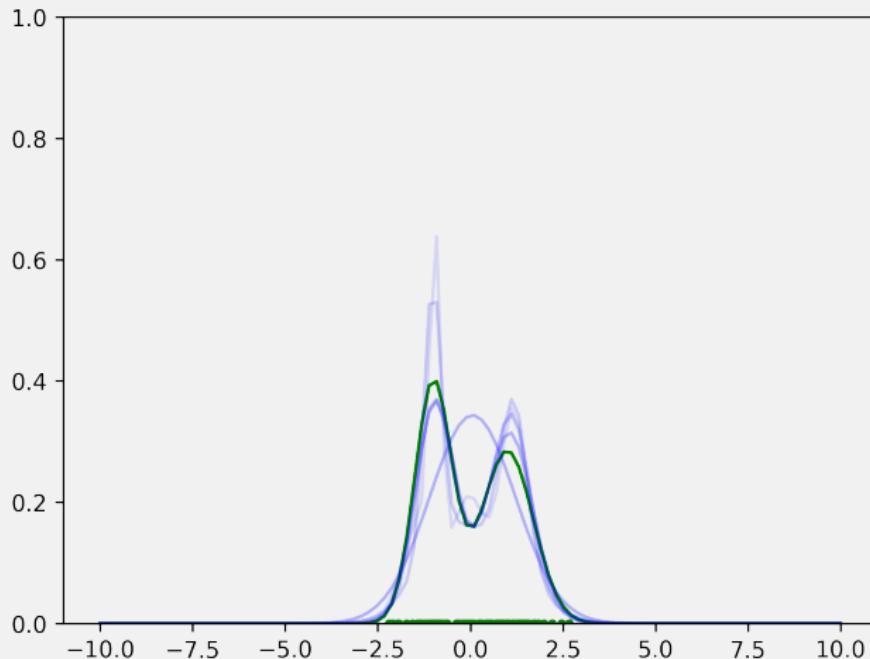
# Bayesi modell-összehasonlítás – demo

n=98, k=1



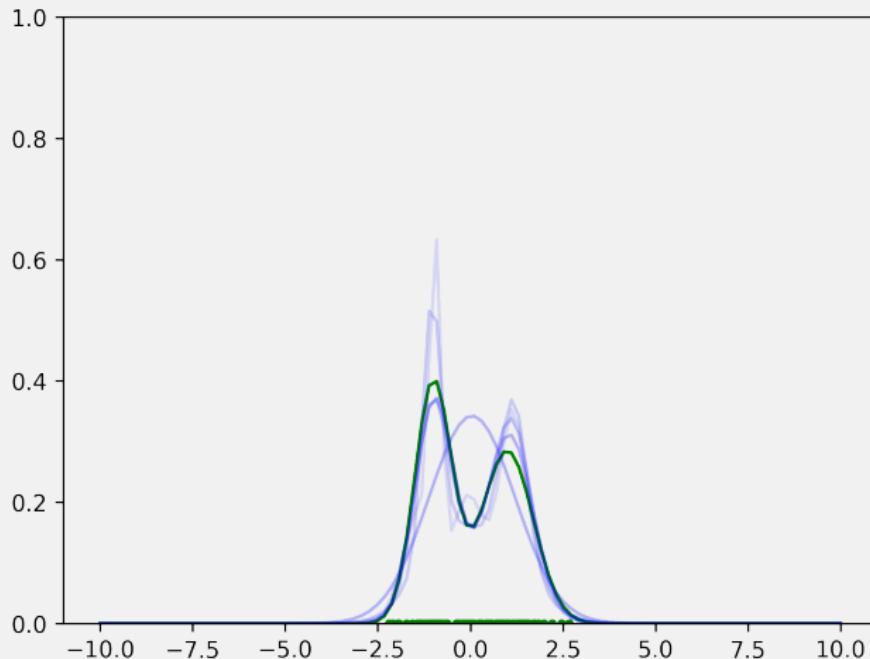
# Bayesi modell-összehasonlítás – demo

n=99, k=1



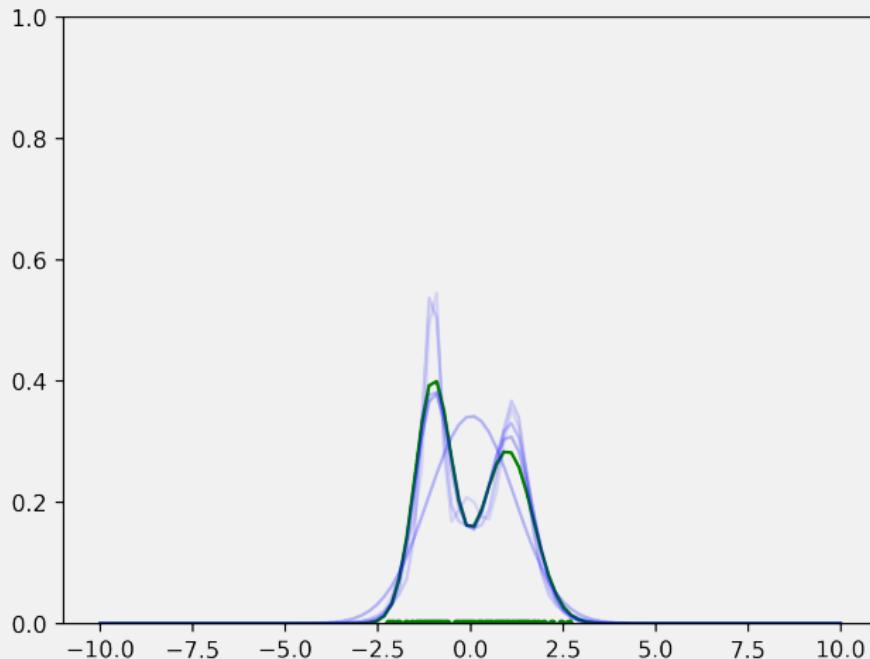
# Bayesi modell-összehasonlítás – demo

$n=100, k=2$



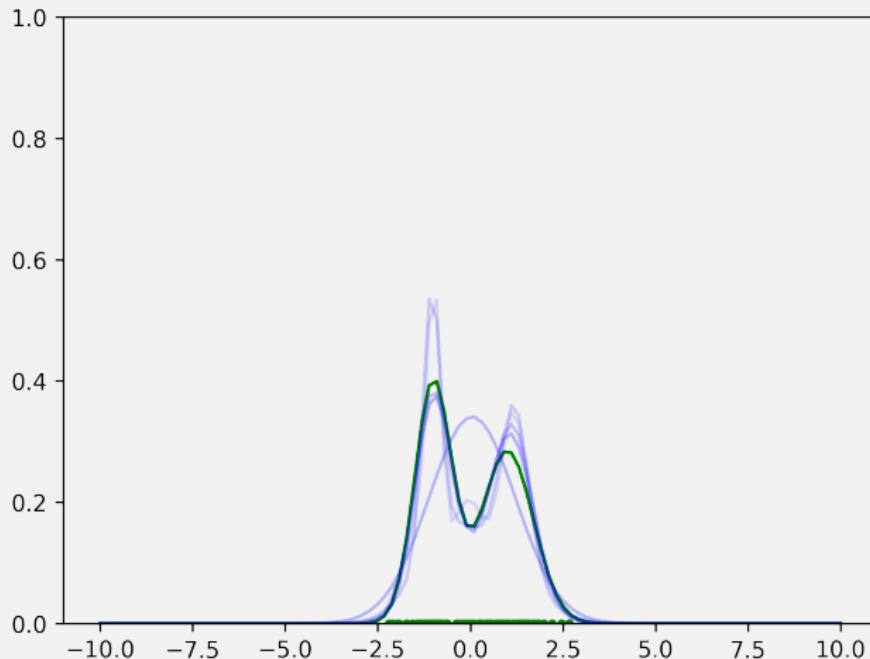
# Bayesi modell-összehasonlítás – demo

$n=101, k=2$



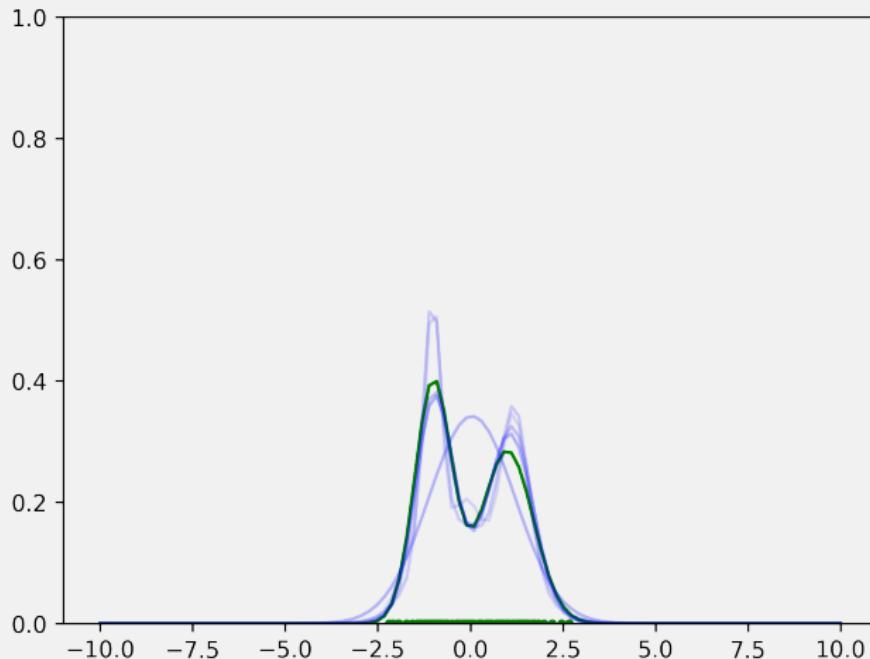
# Bayesi modell-összehasonlítás – demo

$n=102, k=2$



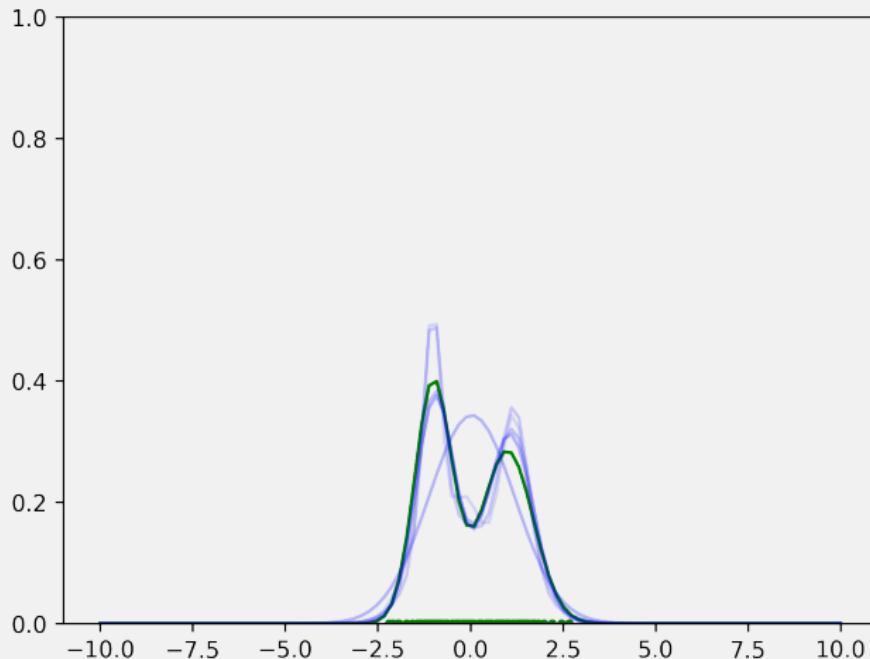
# Bayesi modell-összehasonlítás – demo

$n=103, k=2$



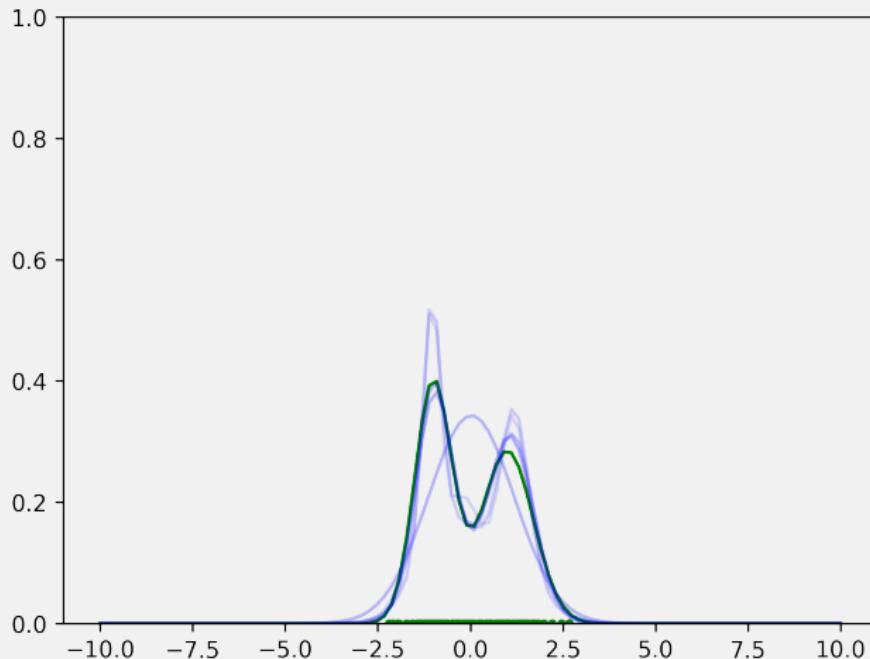
# Bayesi modell-összehasonlítás – demo

$n=104, k=2$



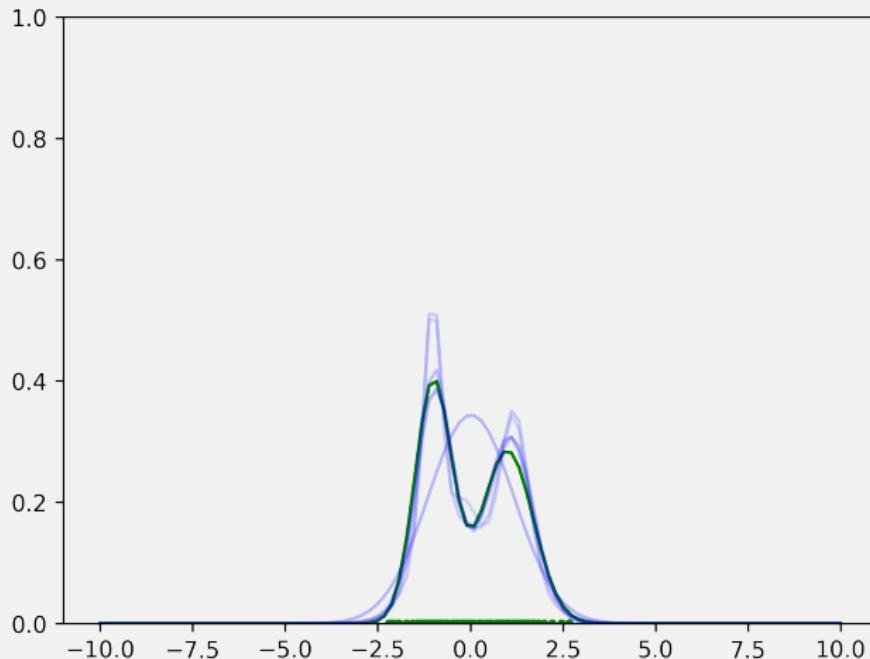
# Bayesi modell-összehasonlítás – demo

$n=105, k=2$



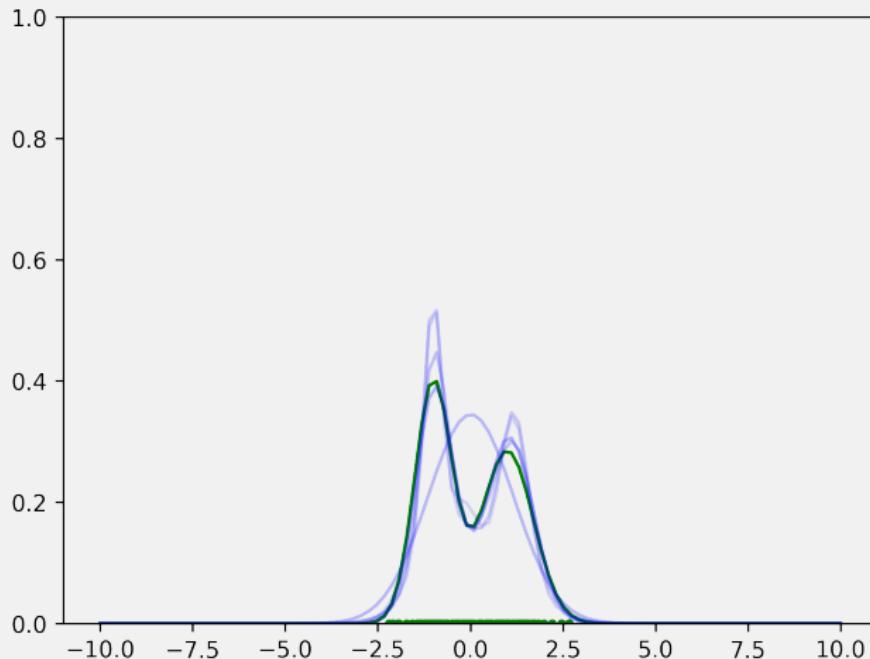
# Bayesi modell-összehasonlítás – demo

$n=106, k=2$



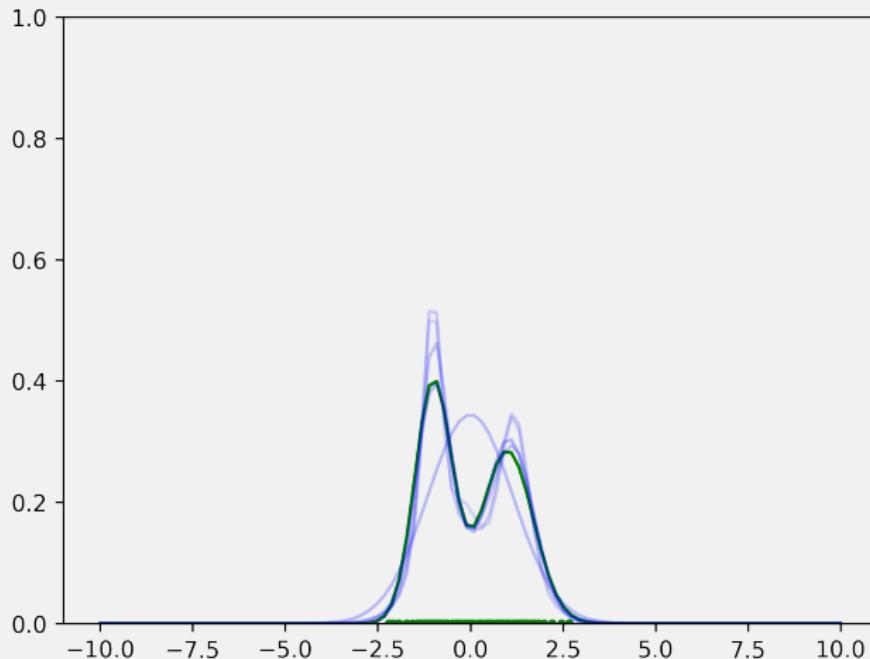
# Bayesi modell-összehasonlítás – demo

$n=107, k=2$



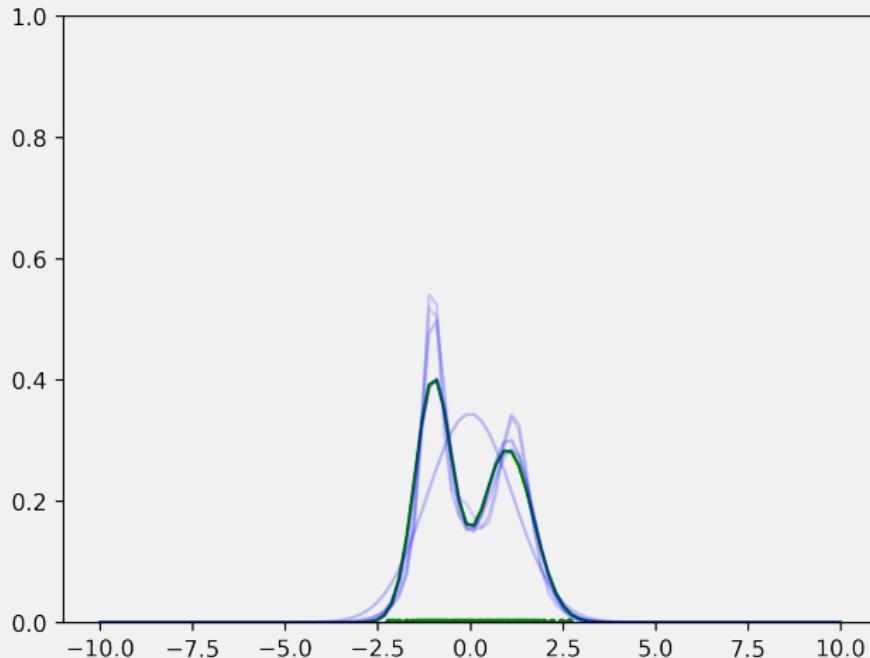
# Bayesi modell-összehasonlítás – demo

$n=108, k=2$



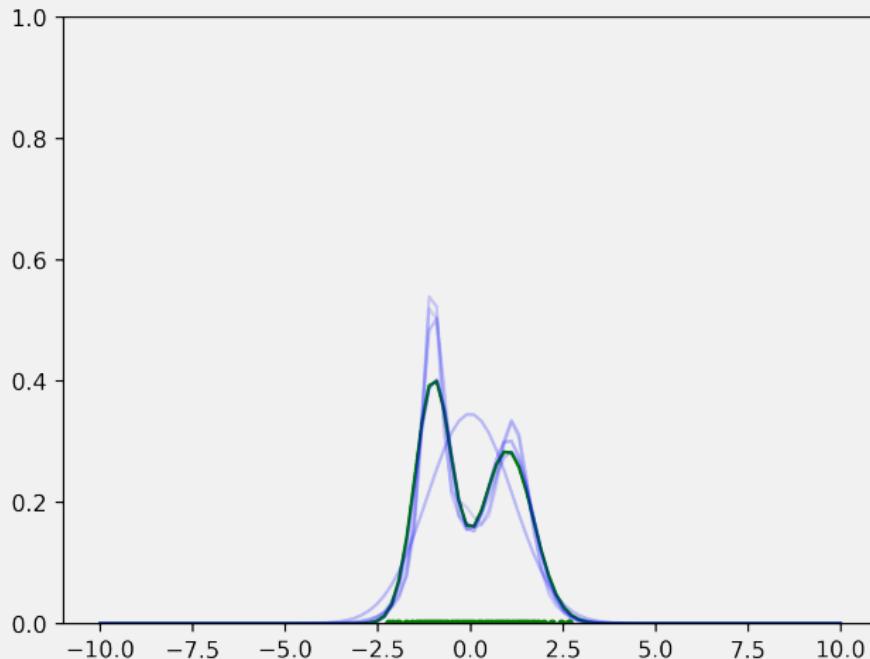
# Bayesi modell-összehasonlítás – demo

$n=109, k=2$



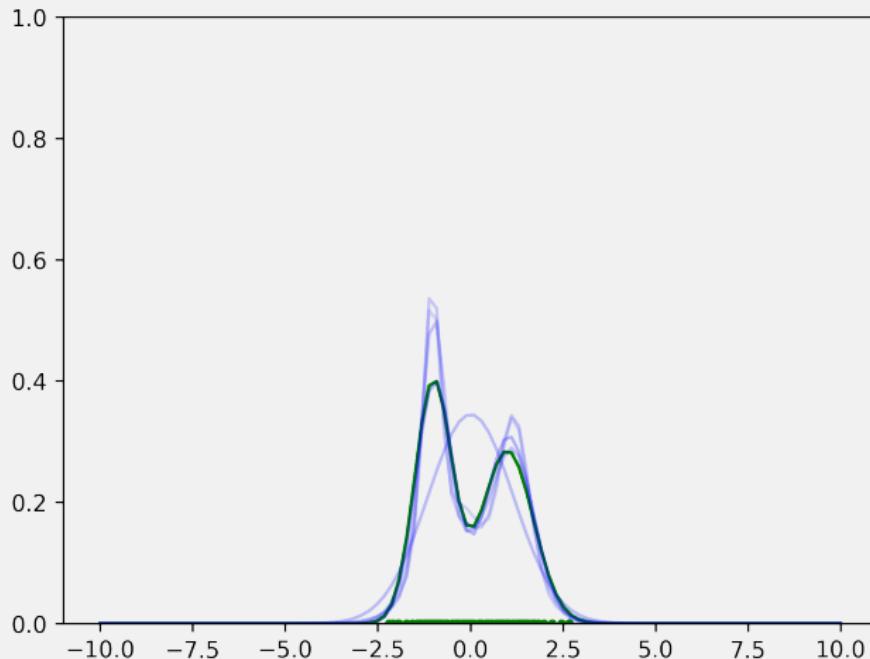
# Bayesi modell-összehasonlítás – demo

$n=110, k=2$



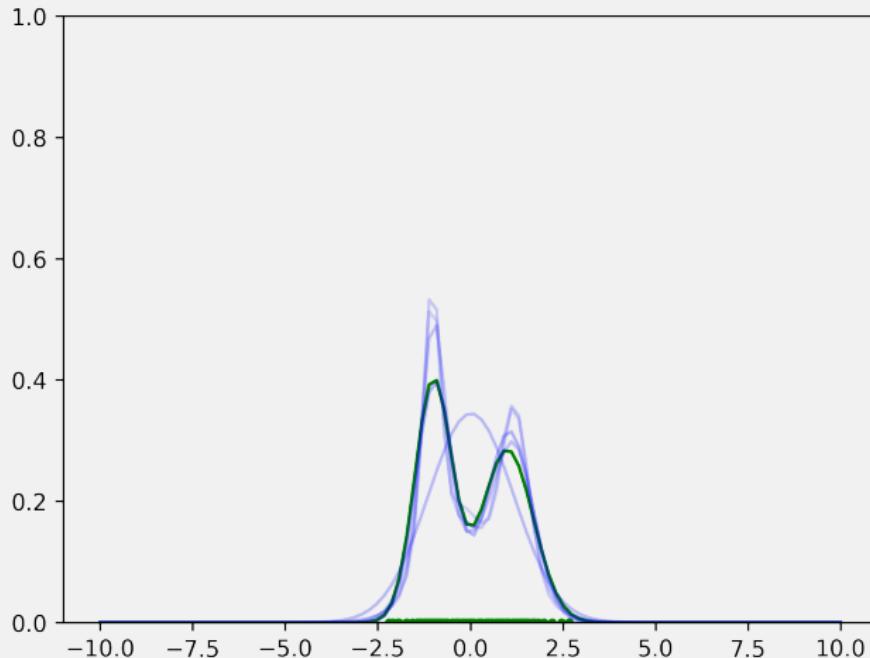
# Bayesi modell-összehasonlítás – demo

$n=111, k=2$



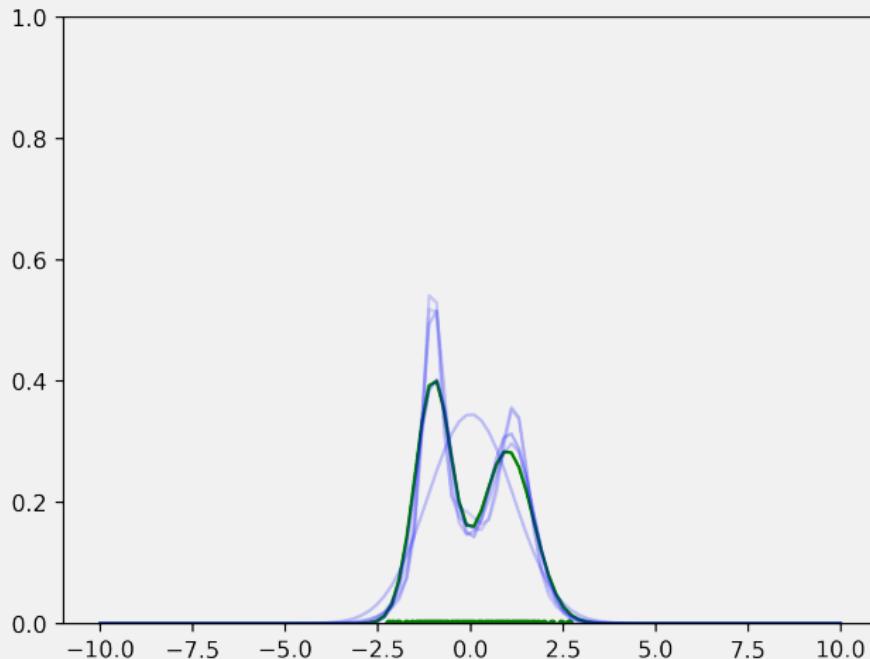
# Bayesi modell-összehasonlítás – demo

$n=112, k=2$



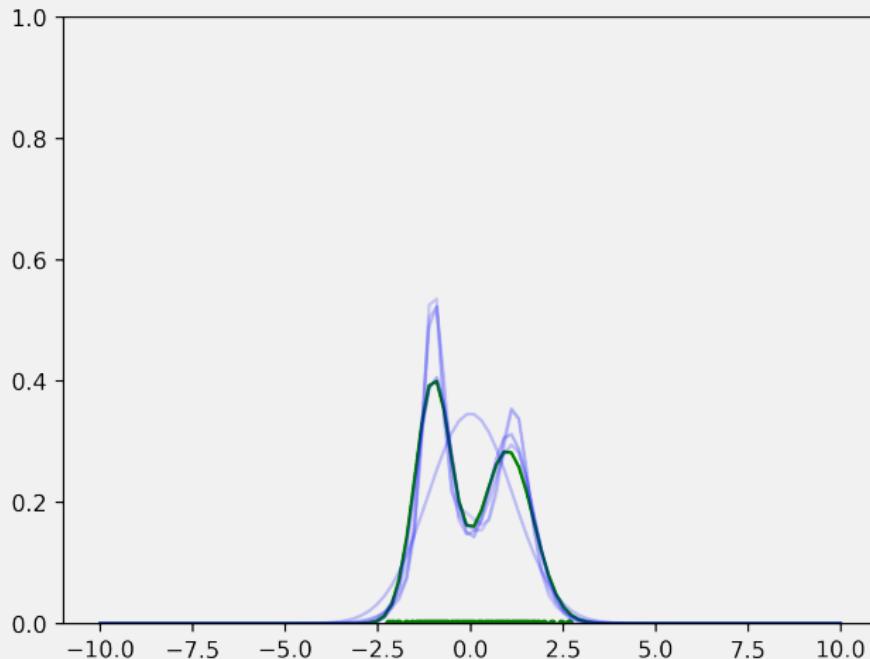
# Bayesi modell-összehasonlítás – demo

$n=113, k=2$



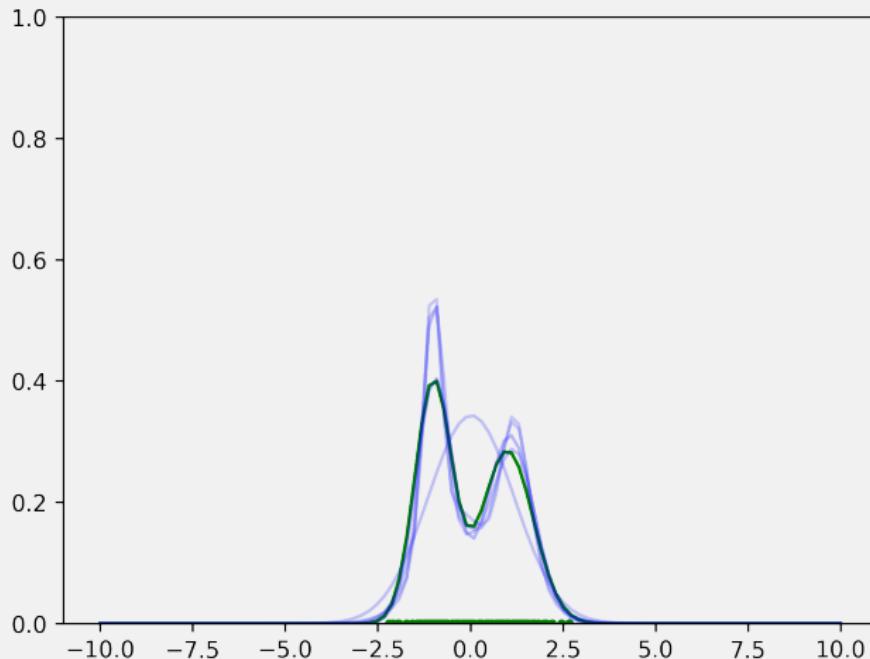
# Bayesi modell-összehasonlítás – demo

$n=114, k=2$



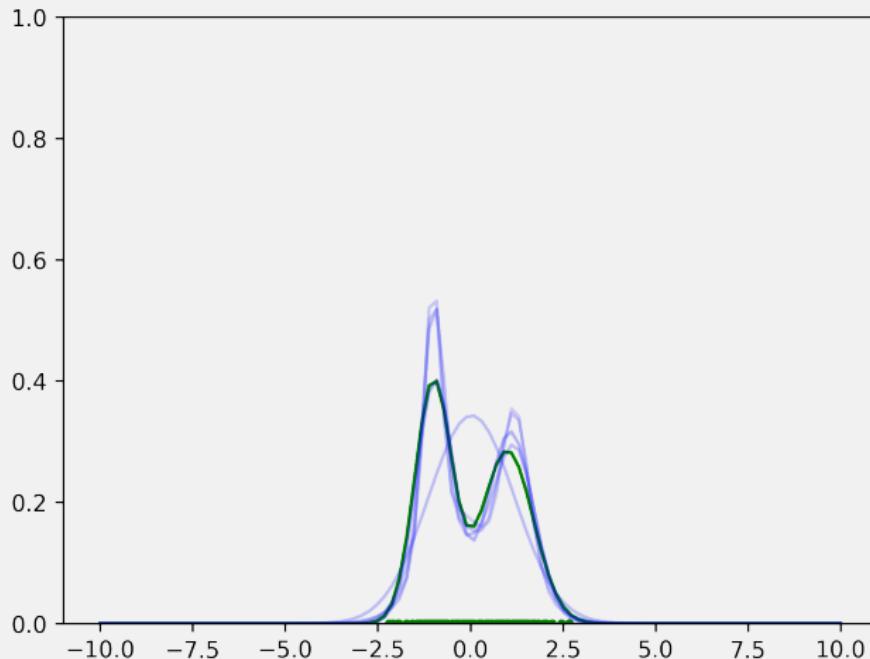
# Bayesi modell-összehasonlítás – demo

$n=115, k=2$



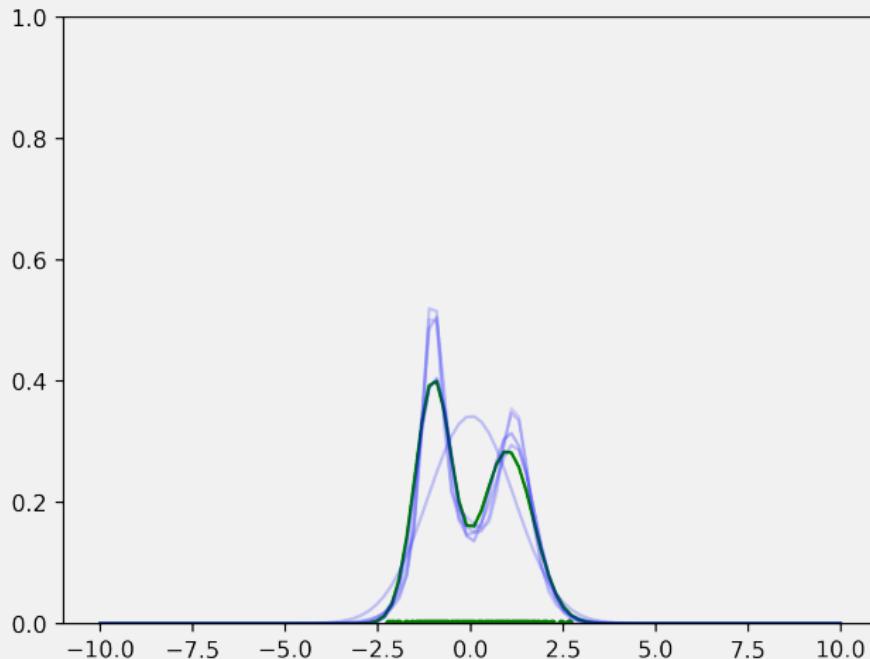
# Bayesi modell-összehasonlítás – demo

$n=116, k=2$



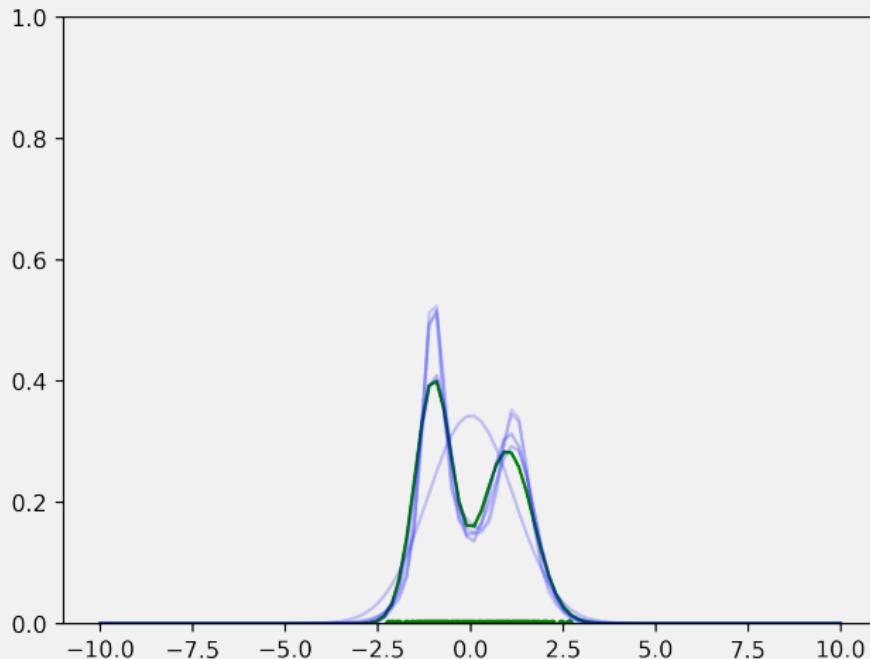
# Bayesi modell-összehasonlítás – demo

$n=117, k=2$



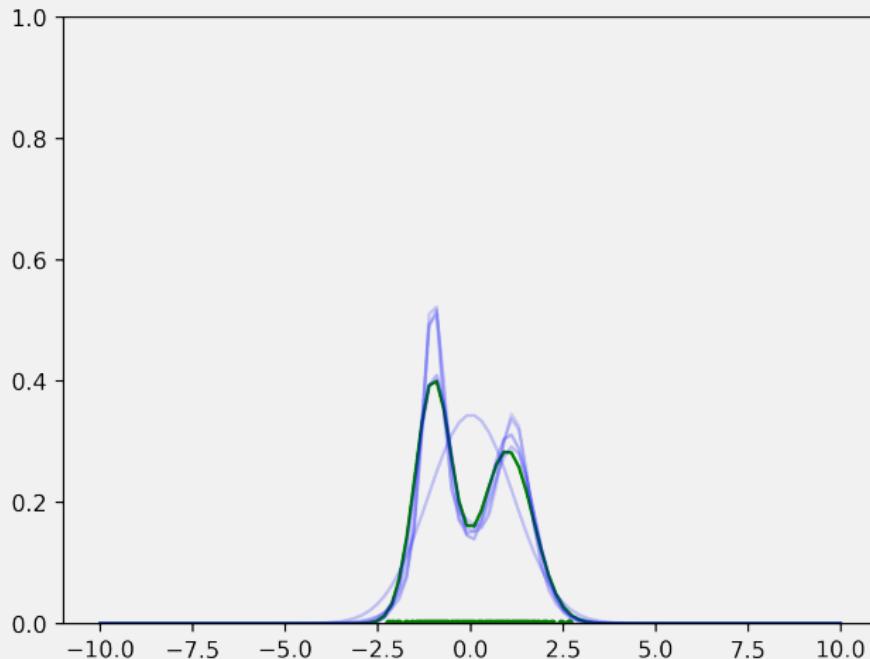
# Bayesi modell-összehasonlítás – demo

$n=118, k=2$



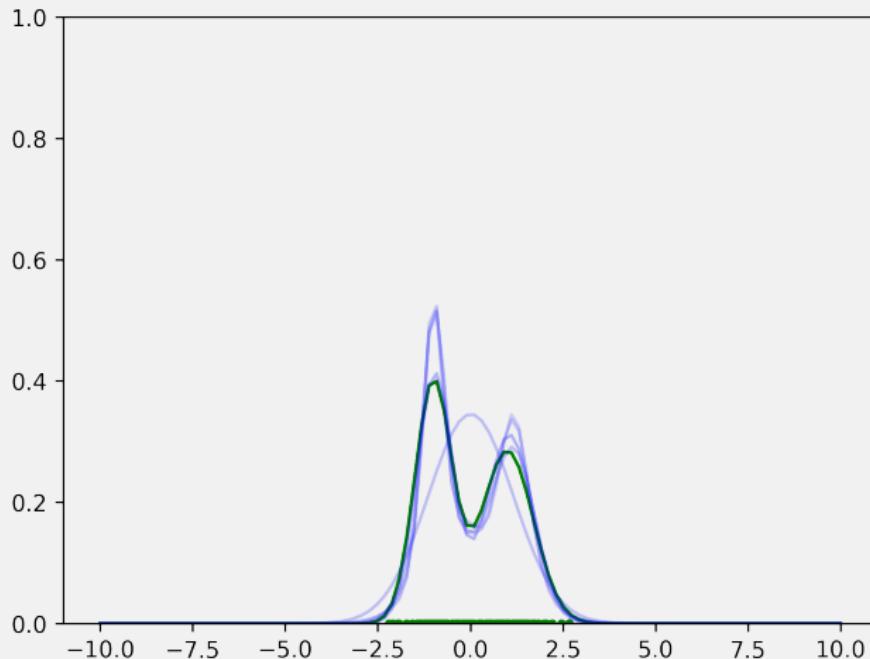
# Bayesi modell-összehasonlítás – demo

$n=119, k=2$



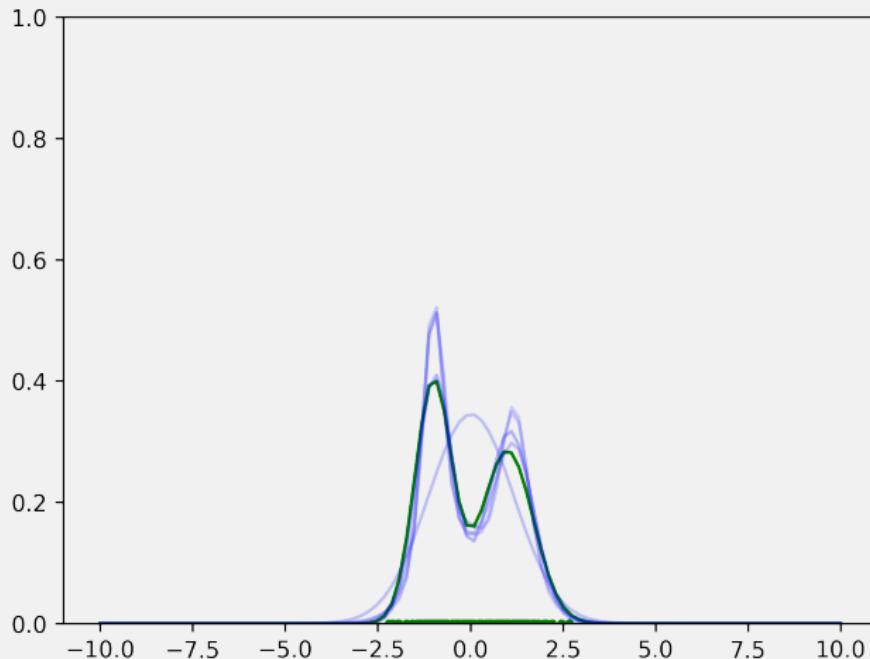
# Bayesi modell-összehasonlítás – demo

$n=120, k=2$



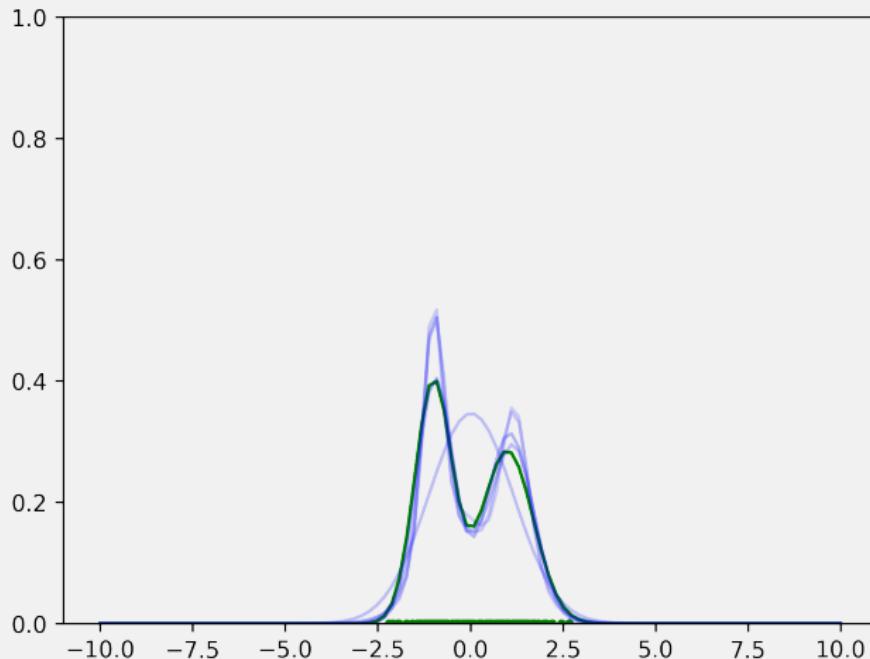
# Bayesi modell-összehasonlítás – demo

$n=121, k=2$



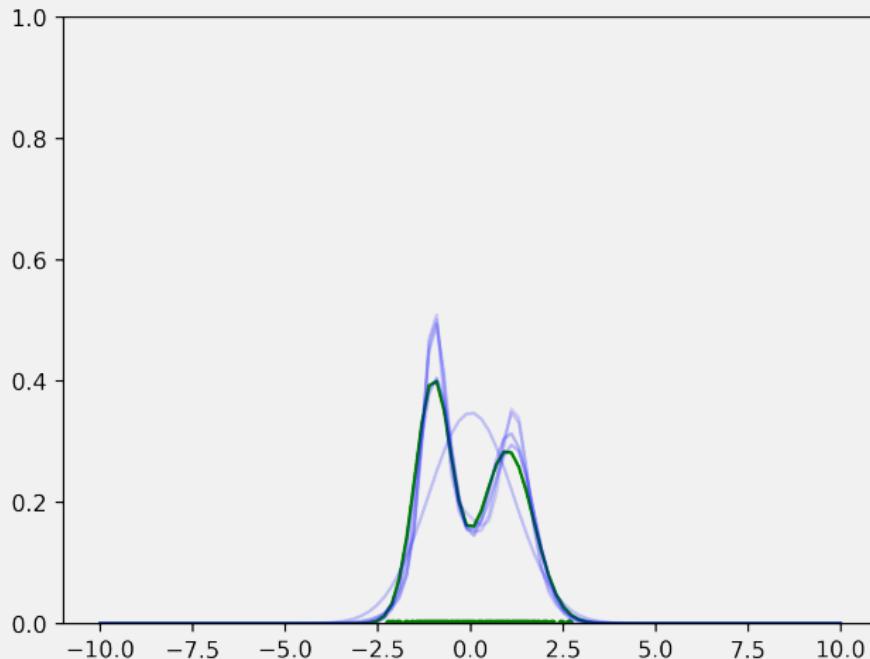
# Bayesi modell-összehasonlítás – demo

$n=122, k=2$



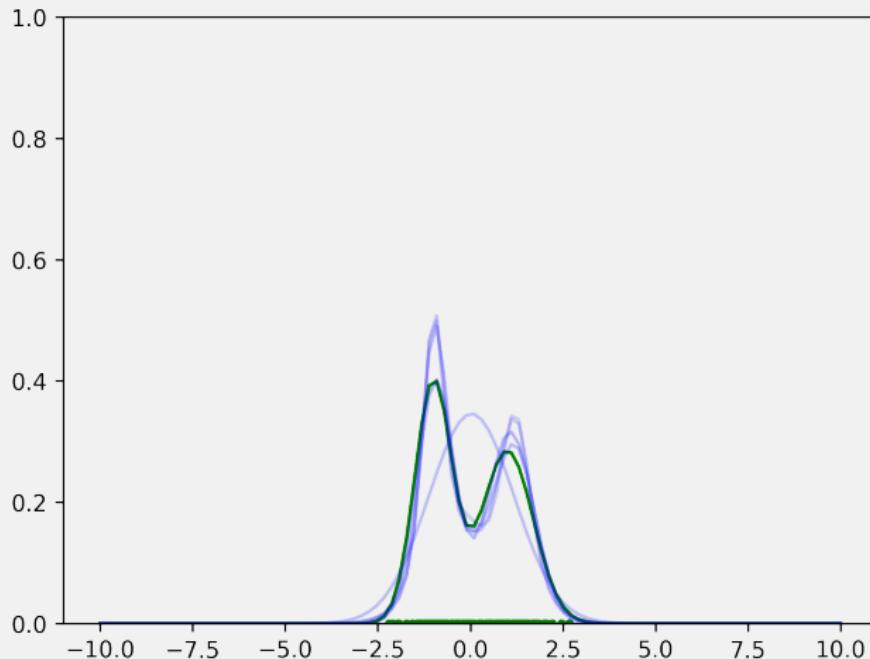
# Bayesi modell-összehasonlítás – demo

$n=123, k=2$



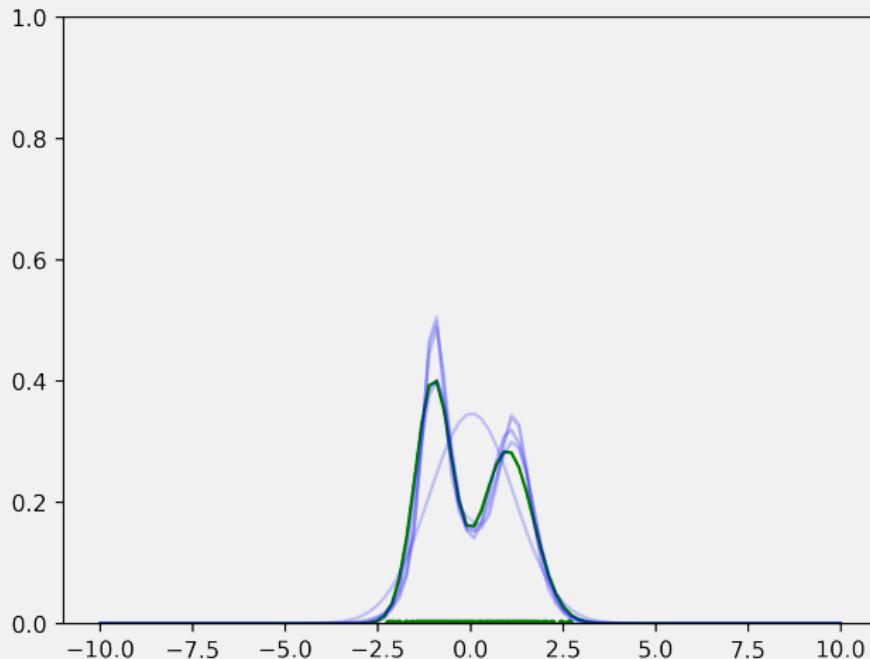
# Bayesi modell-összehasonlítás – demo

$n=124, k=2$



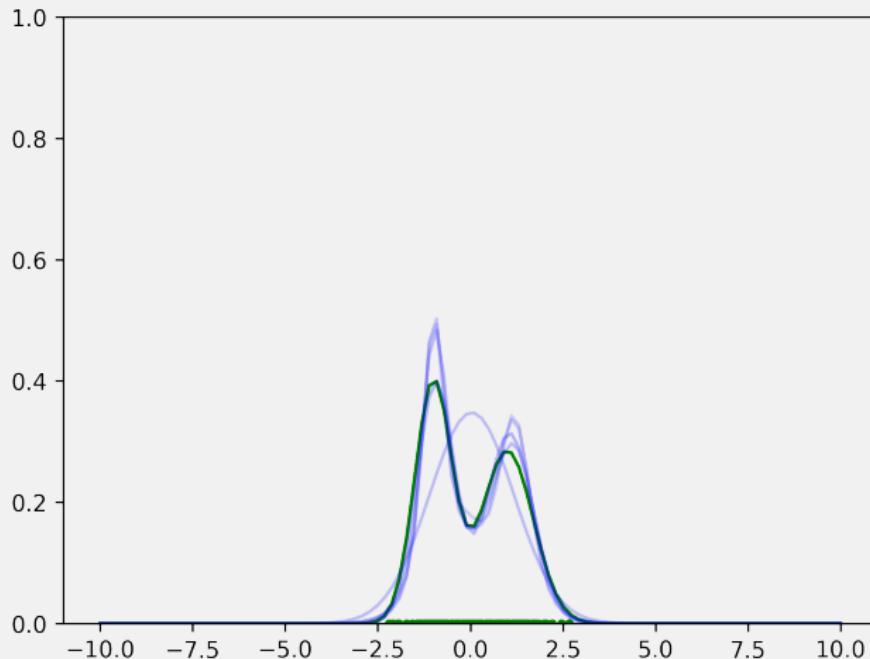
# Bayesi modell-összehasonlítás – demo

$n=125, k=2$



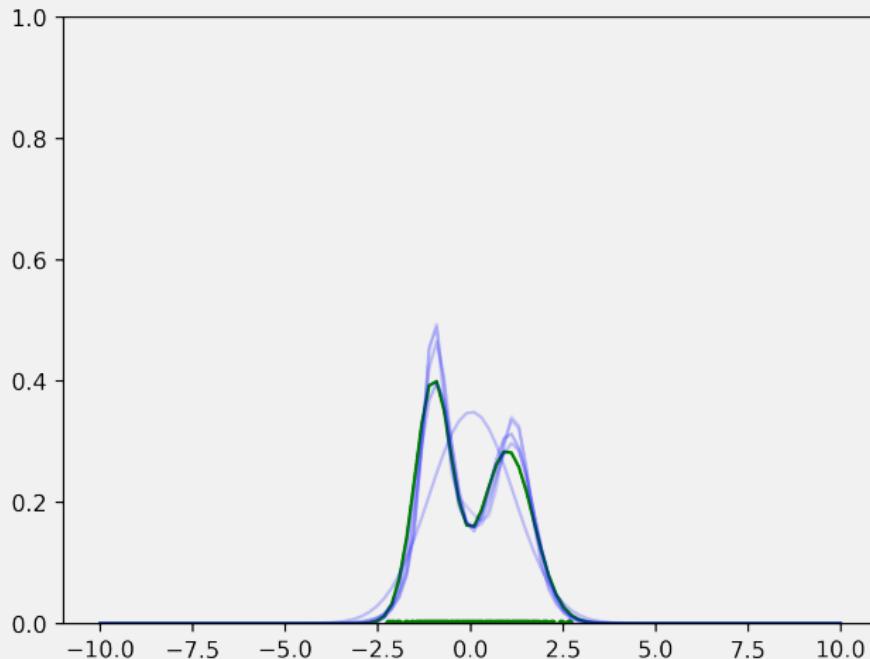
# Bayesi modell-összehasonlítás – demo

$n=126, k=2$



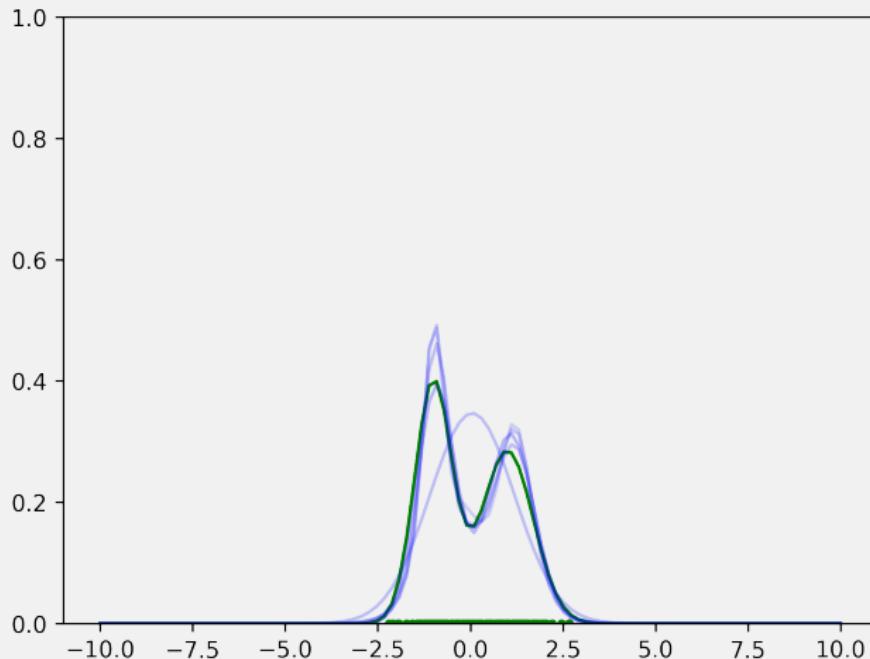
# Bayesi modell-összehasonlítás – demo

$n=127, k=2$



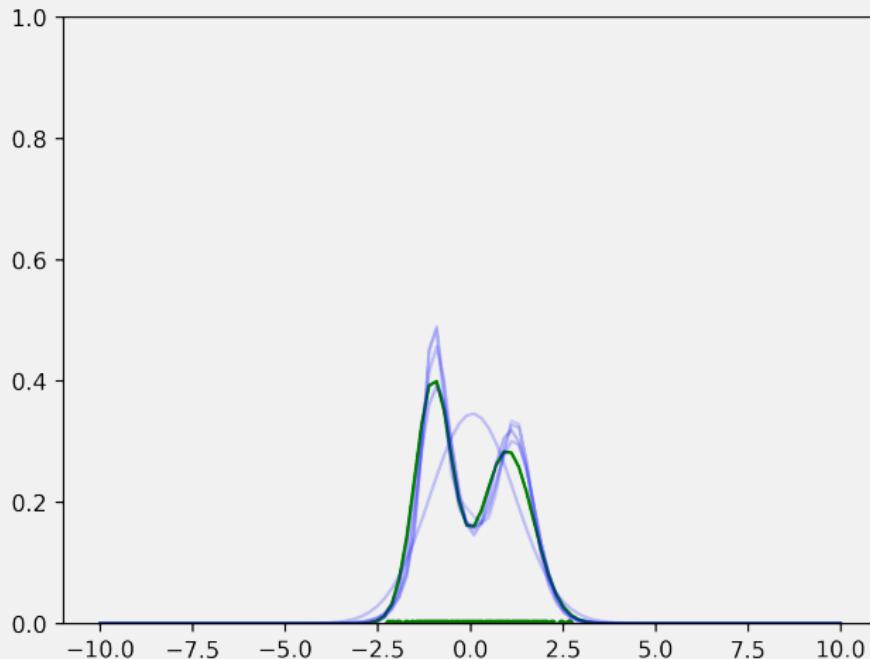
# Bayesi modell-összehasonlítás – demo

$n=128, k=2$



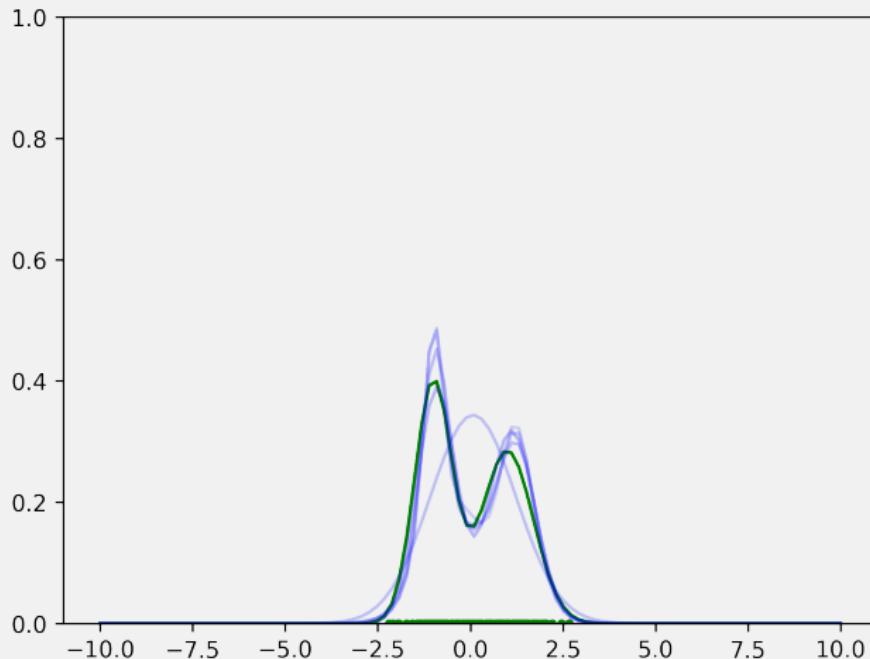
# Bayesi modell-összehasonlítás – demo

$n=129, k=2$



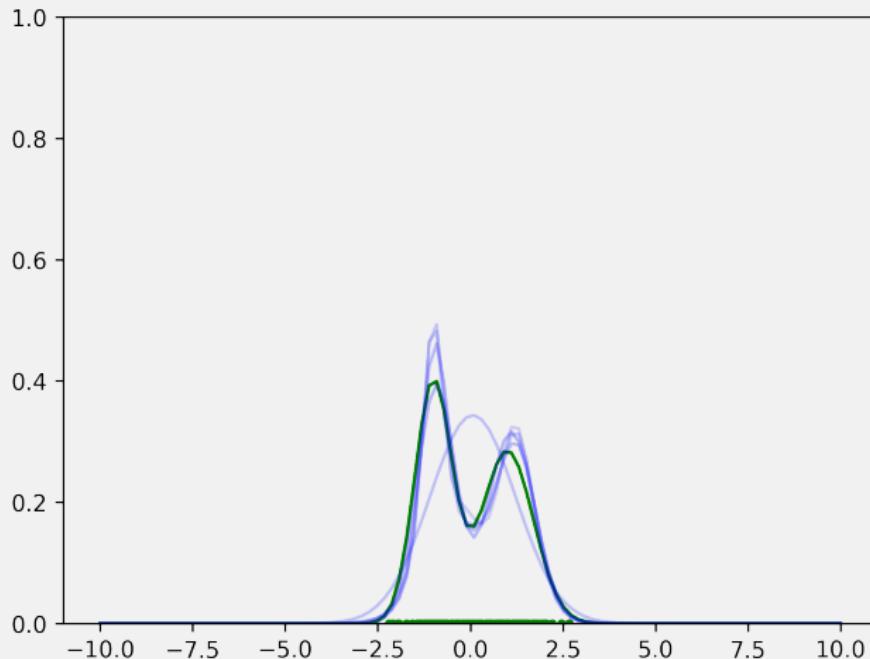
# Bayesi modell-összehasonlítás – demo

$n=130, k=2$



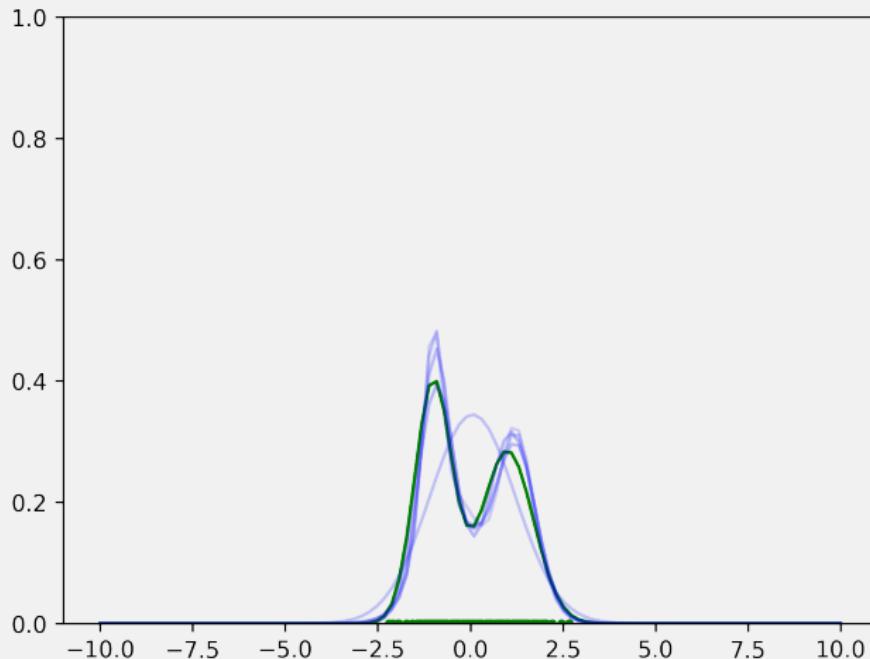
# Bayesi modell-összehasonlítás – demo

$n=131, k=2$



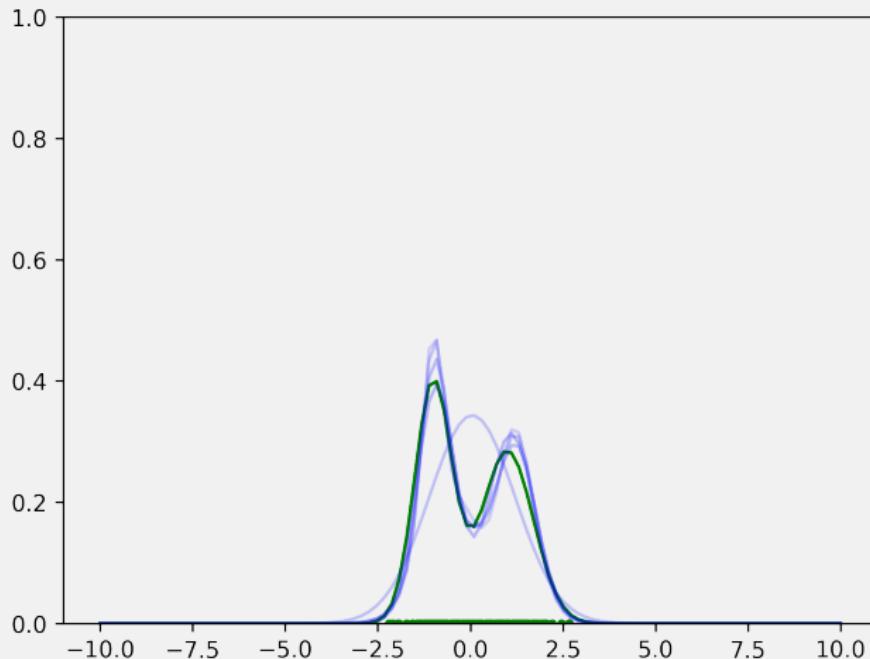
# Bayesi modell-összehasonlítás – demo

$n=132, k=2$



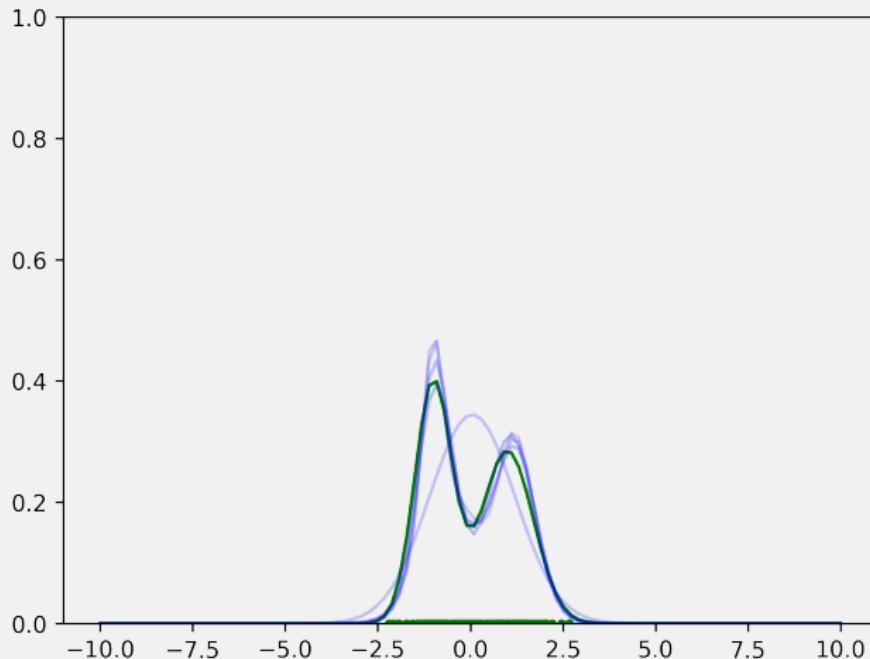
# Bayesi modell-összehasonlítás – demo

$n=133, k=2$



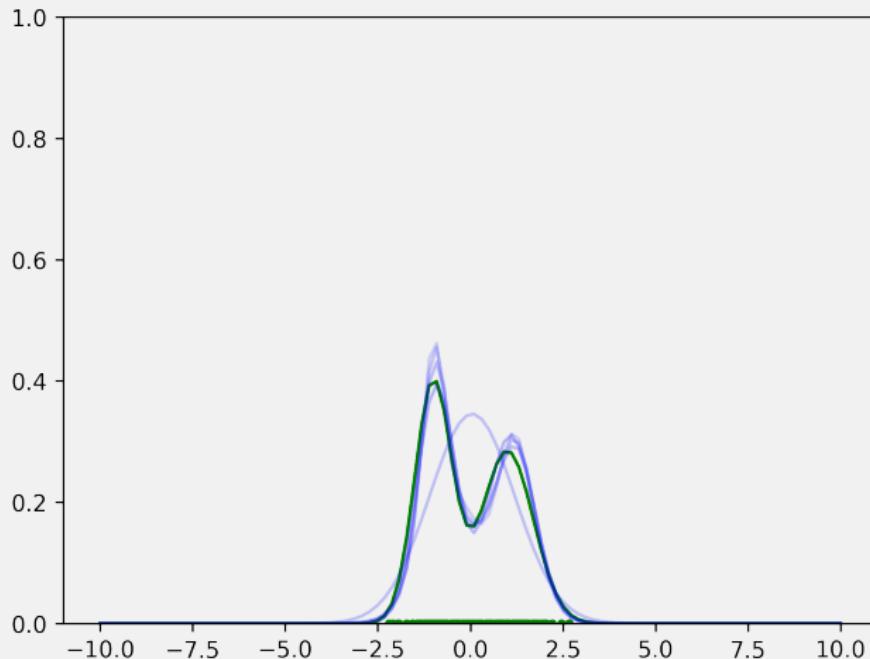
# Bayesi modell-összehasonlítás – demo

$n=134, k=2$



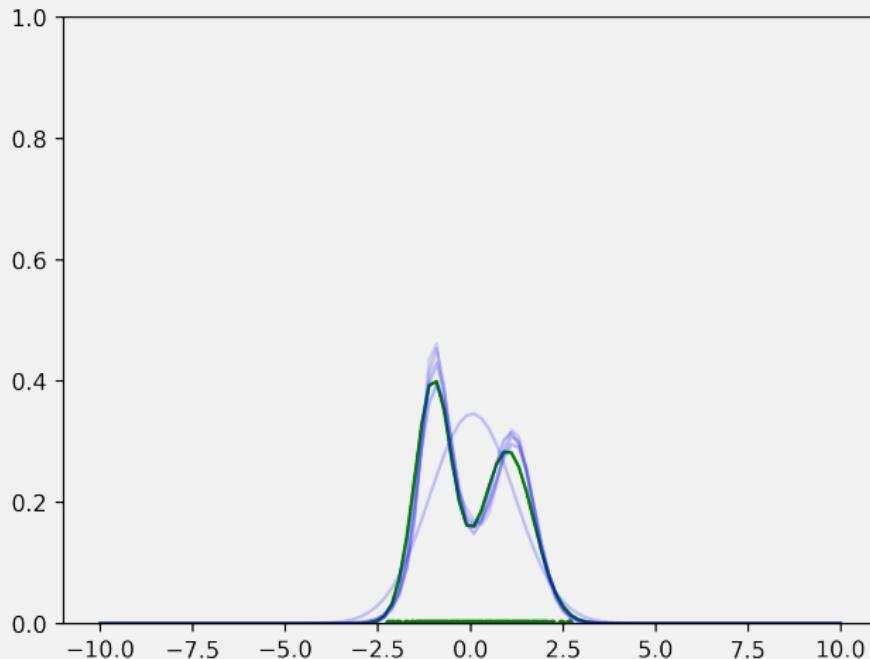
# Bayesi modell-összehasonlítás – demo

$n=135, k=2$



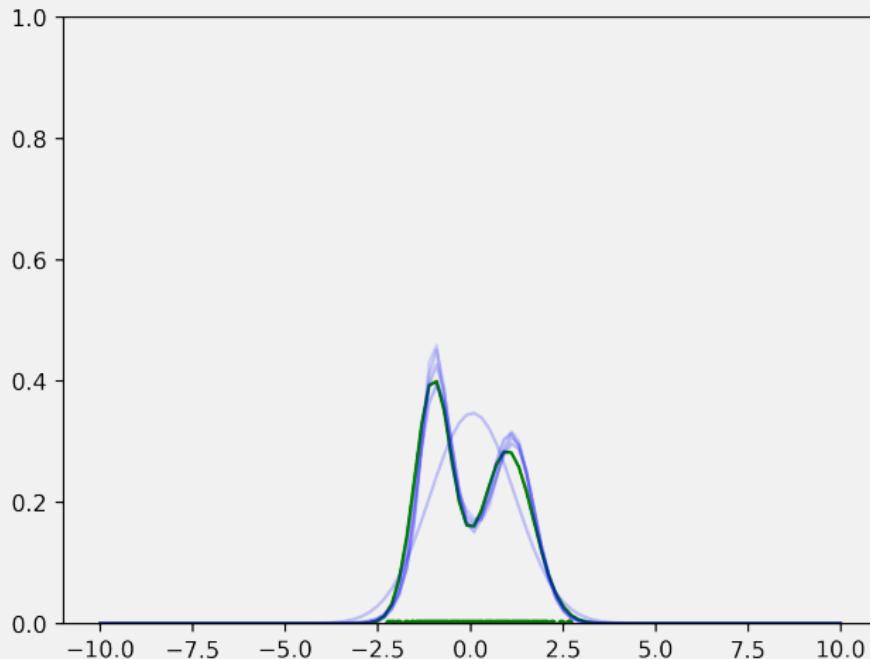
# Bayesi modell-összehasonlítás – demo

$n=136, k=2$



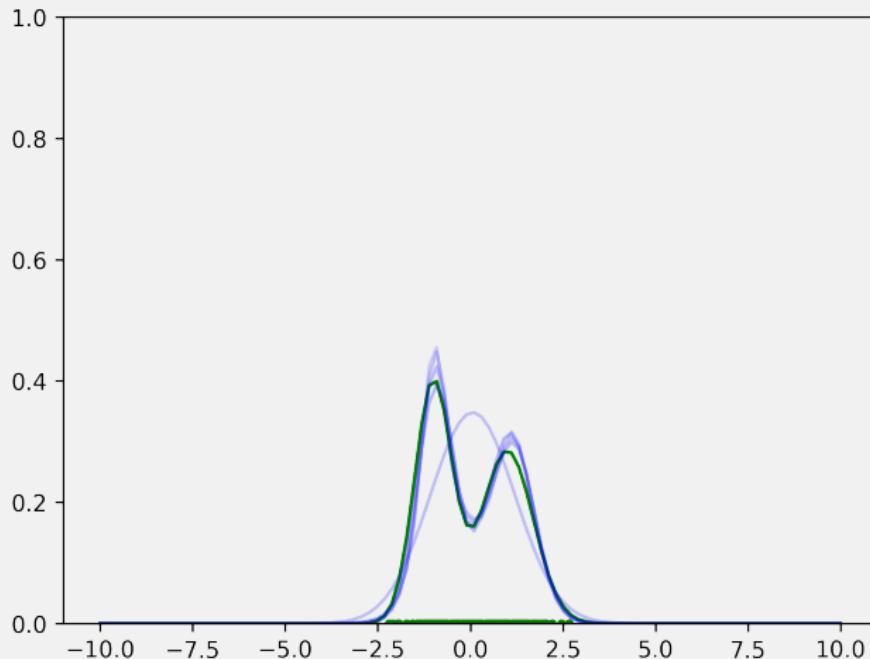
# Bayesi modell-összehasonlítás – demo

$n=137, k=2$



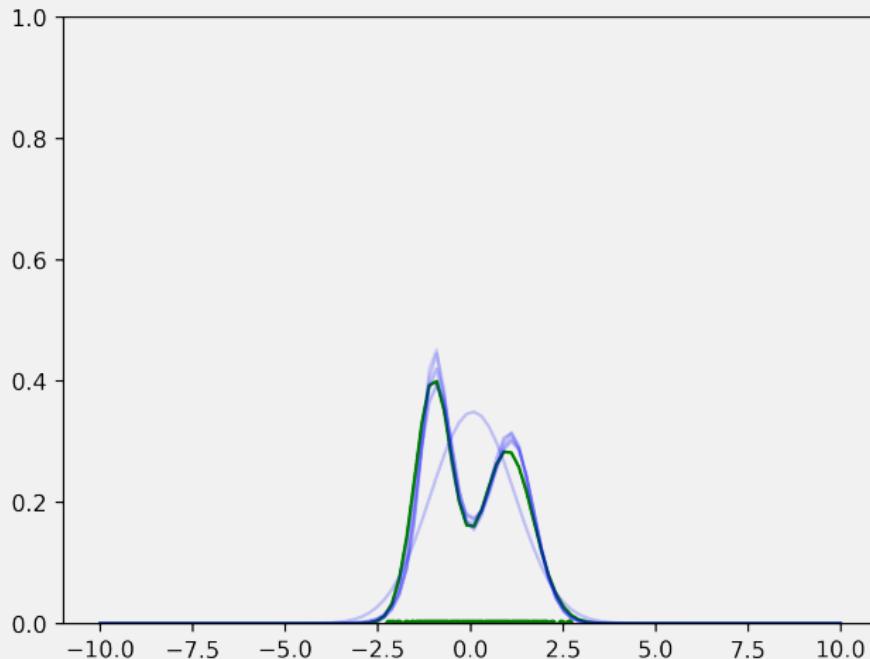
# Bayesi modell-összehasonlítás – demo

$n=138, k=2$



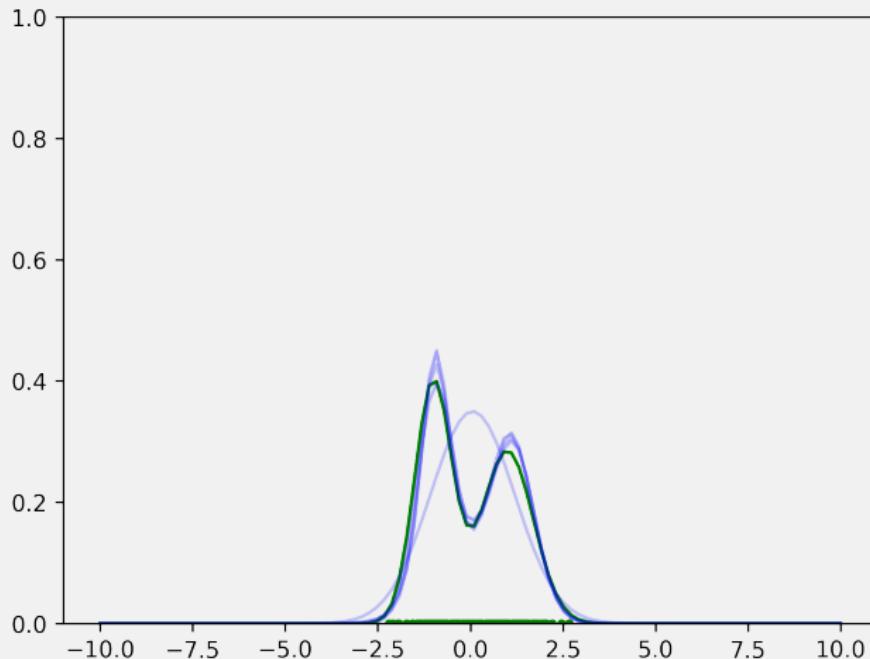
# Bayesi modell-összehasonlítás – demo

$n=139, k=2$



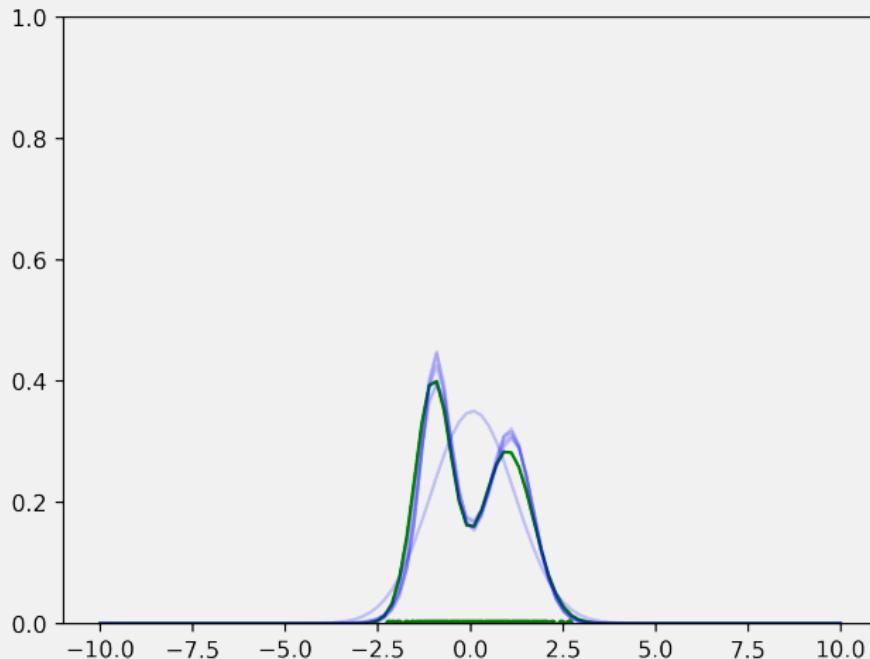
# Bayesi modell-összehasonlítás – demo

$n=140, k=2$



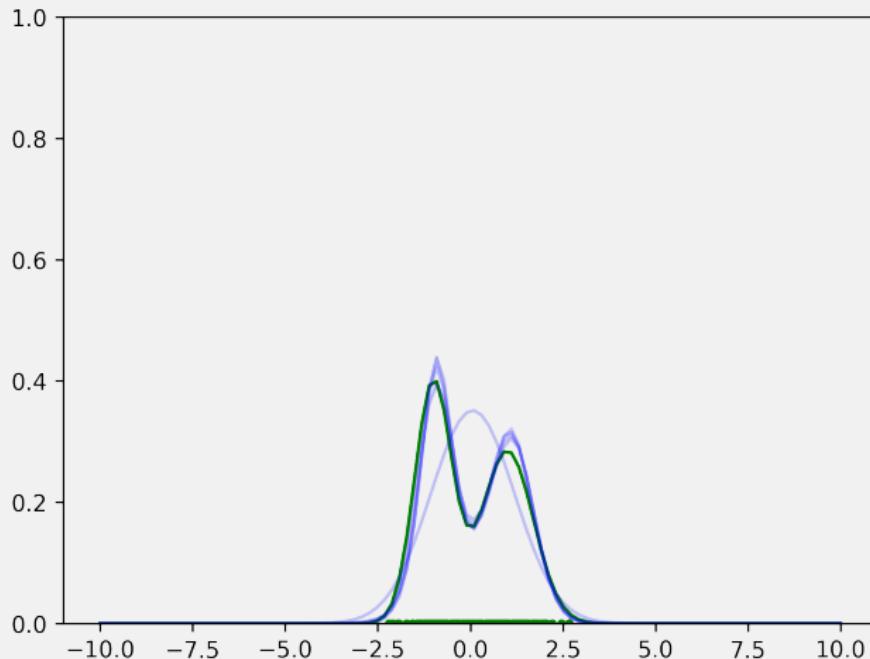
# Bayesi modell-összehasonlítás – demo

$n=141, k=2$



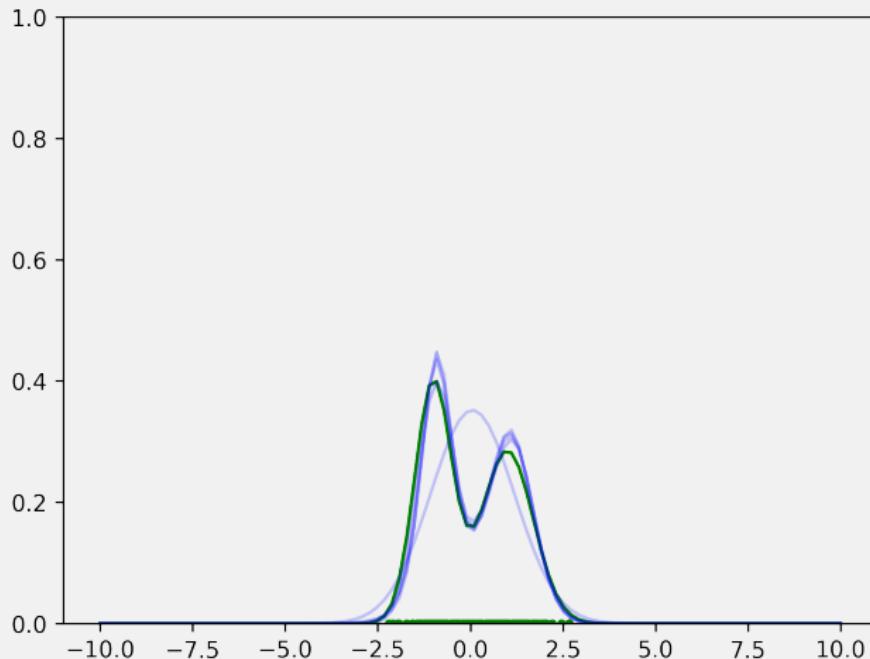
# Bayesi modell-összehasonlítás – demo

$n=142, k=2$



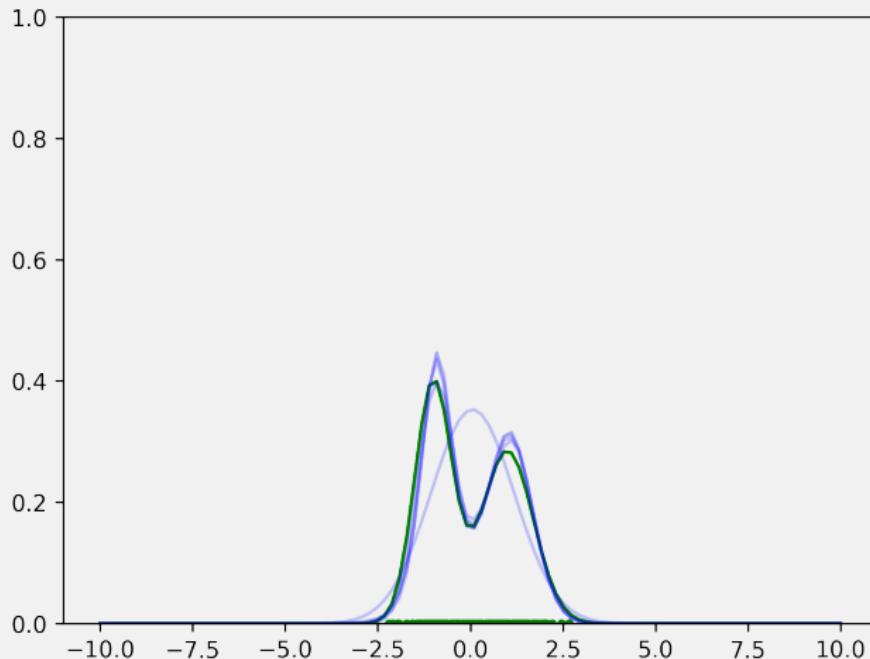
# Bayesi modell-összehasonlítás – demo

$n=143, k=2$



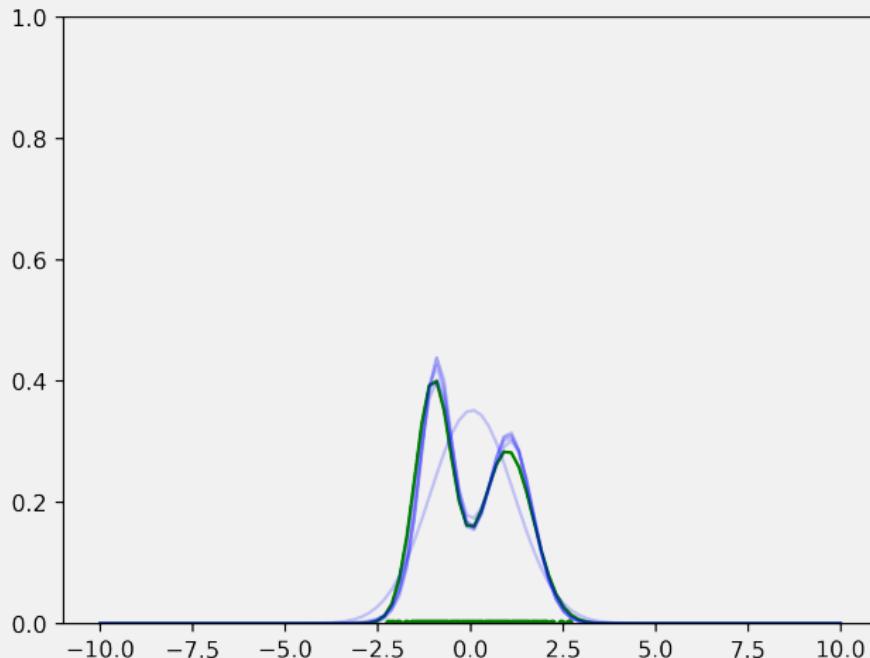
# Bayesi modell-összehasonlítás – demo

$n=144, k=2$



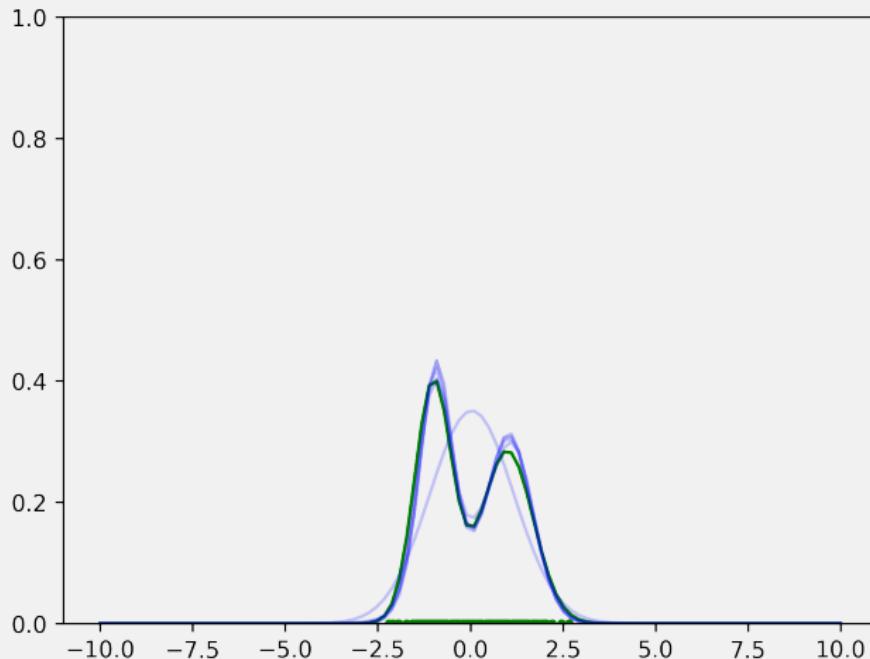
# Bayesi modell-összehasonlítás – demo

$n=145, k=2$



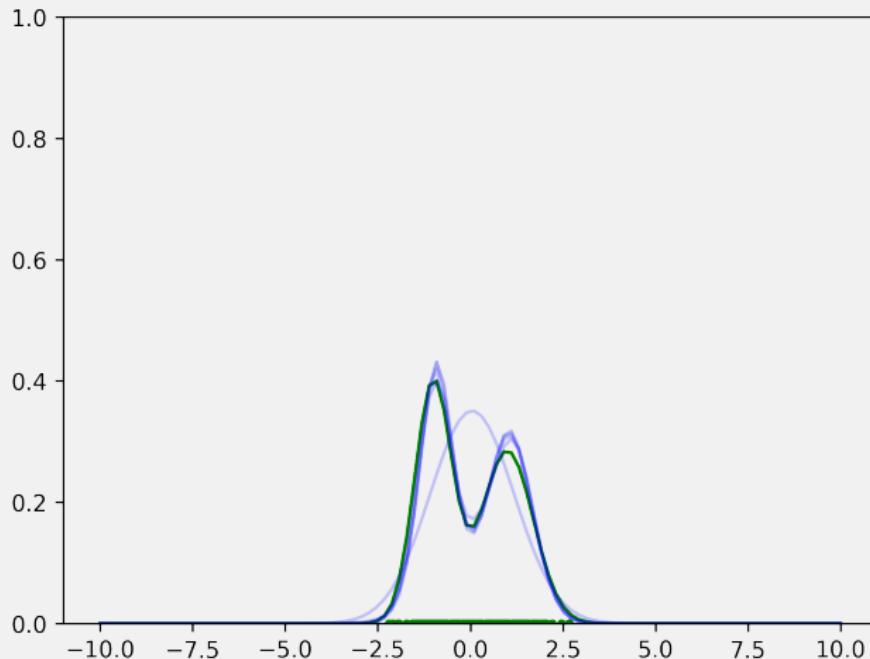
# Bayesi modell-összehasonlítás – demo

$n=146, k=2$



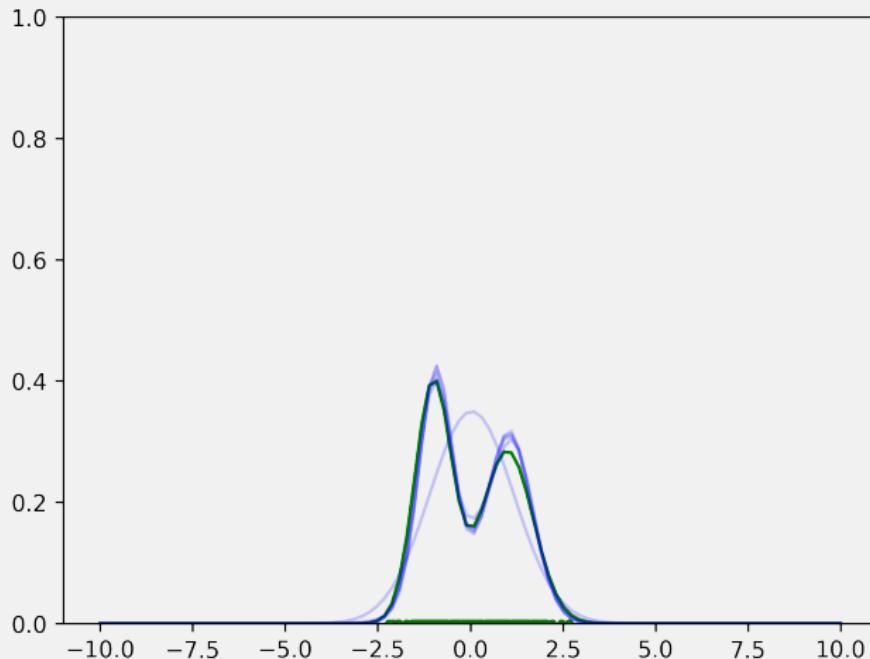
# Bayesi modell-összehasonlítás – demo

$n=147, k=2$



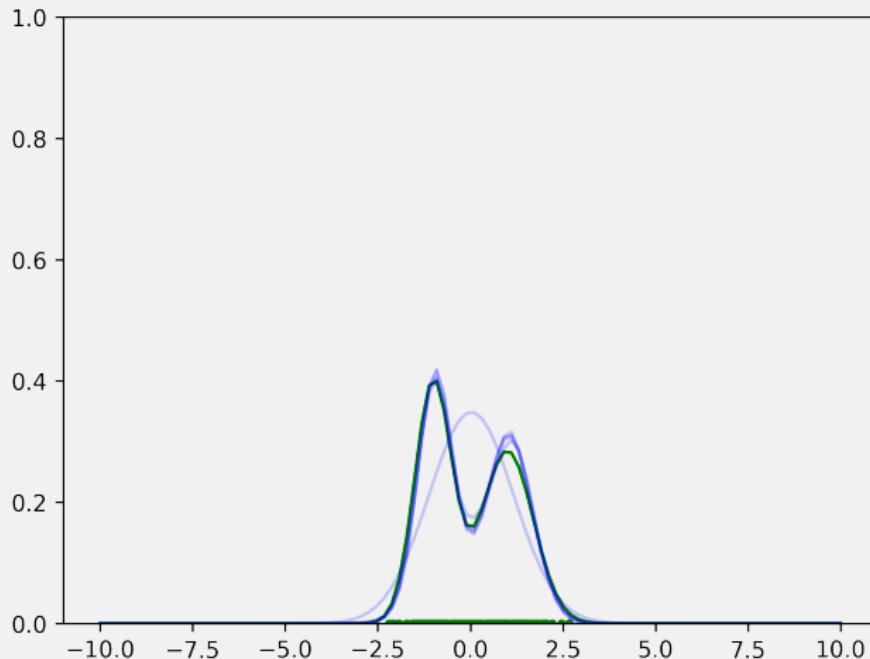
# Bayesi modell-összehasonlítás – demo

$n=148, k=2$



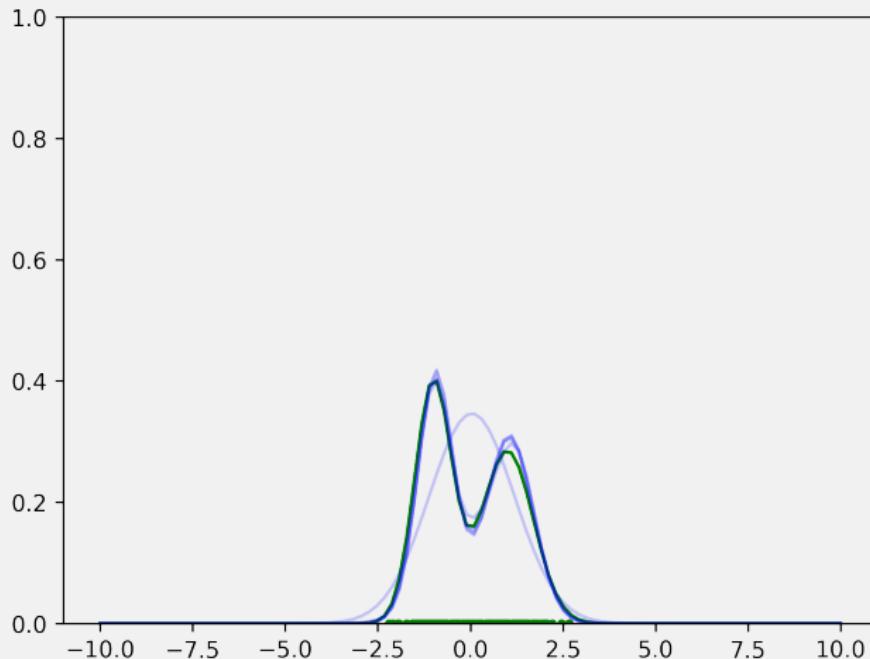
# Bayesi modell-összehasonlítás – demo

$n=149, k=2$

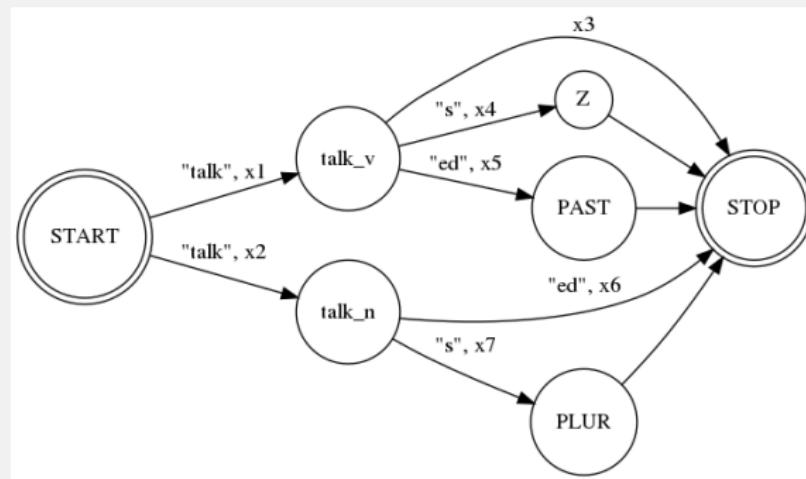


# Bayesi modell-összehasonlítás – demo

$n=150, k=2$

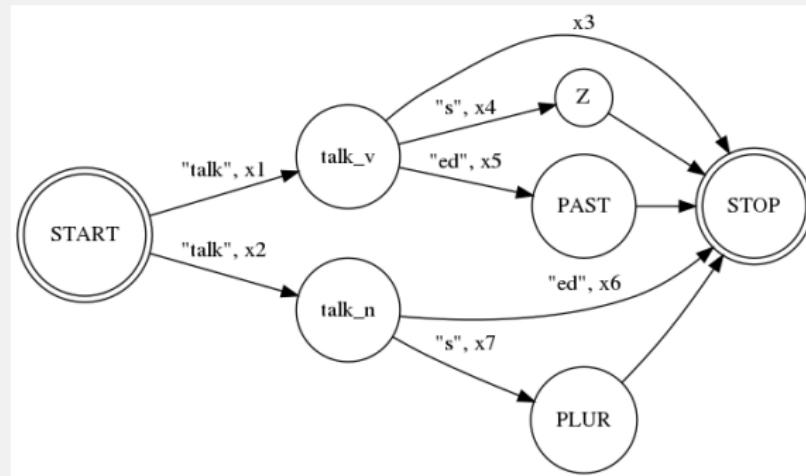


# Véges állapotú morfológia



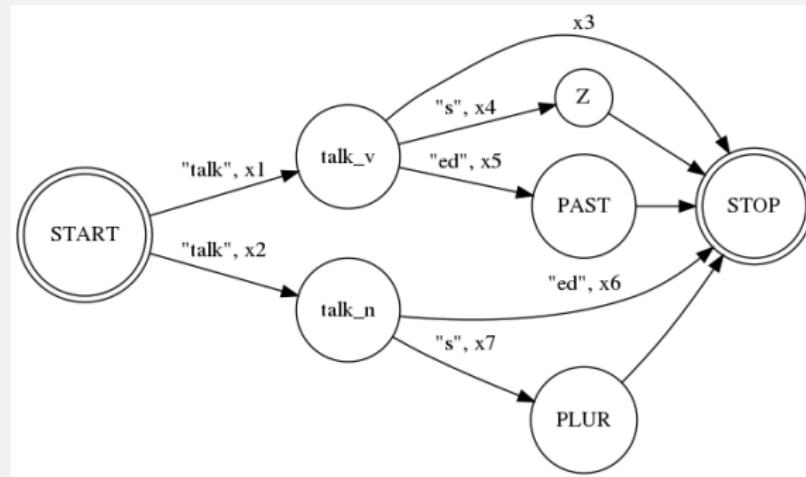
- Egy eloszlást definiál a  $\{talk, talks, talked\}$  halmaz felett.

# Véges állapotú morfológia



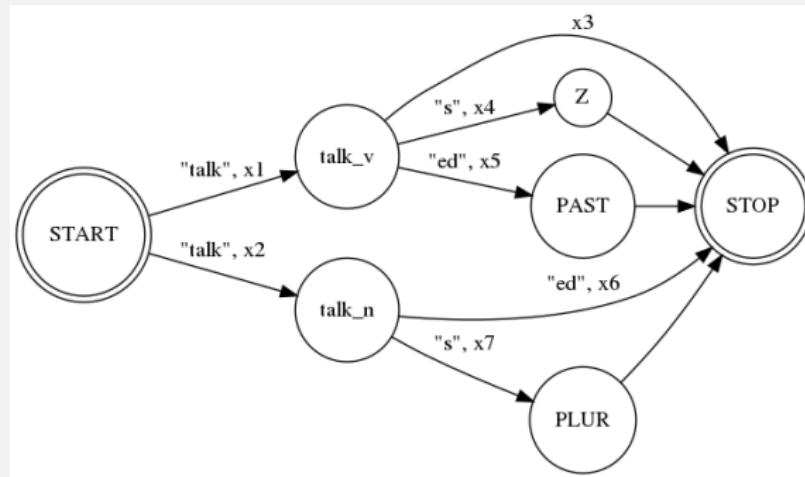
- Egy eloszlást definiál a  $\{talk, talks, talked\}$  halmaz felett.
- az eloszlás differenciálható módon függ az  $x_1, x_2 \dots x_7$  paraméterektől.

# Véges állapotú morfológia



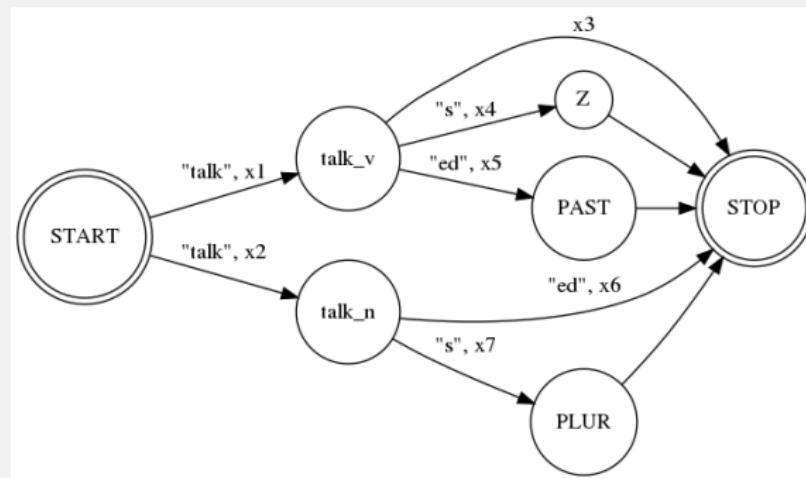
- Egy eloszlást definiál a  $\{talk, talks, talked\}$  halmaz felett.
- az eloszlás differenciálható módon függ az  $x_1, x_2 \dots x_7$  paraméterektől.
- A paraméterekre megkötések vonatkoznak:

# Véges állapotú morfológia



- Egy eloszlást definiál a  $\{talk, talks, talked\}$  halmaz felett.
- az eloszlás differenciálható módon függ az  $x_1, x_2 \dots x_7$  paraméterektől.
- A paraméterekre megkötések vonatkoznak:
  - $x_i > 0$

# Véges állapotú morfológia



- Egy eloszlást definiál a  $\{talk, talks, talked\}$  halmaz felett.
- az eloszlás differenciálható módon függ az  $x_1, x_2 \dots x_7$  paraméterektől.
- A paraméterekre megkötések vonatkoznak:
  - $x_i > 0$
  - $x_1 + x_2 = x_3 + x_4 + x_5 = x_6 + x_7 = 1$

# Véges állapotú automata tanítása

- Mi legyen az architektúra?

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat
- Adott architektúrára az élsúlyok megtalálása.

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat
- Adott architektúrára az élsúlyok megtalálása.
  - folytonos optimalizálási feladat

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat
- Adott architektúrára az élsúlyok megtalálása.
  - folytonos optimalizálási feladat
  - constrained optimization

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat
- Adott architektúrára az élsúlyok megtalálása.
  - folytonos optimalizálási feladat
  - constrained optimization
- Keressük az optimális automatát genetikus algoritmussal  
(Csikós, 2019)

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat
- Adott architektúrára az élsúlyok megtalálása.
  - folytonos optimalizálási feladat
  - constrained optimization
- Keressük az optimális automatát genetikus algoritmussal  
(Csikós, 2019)
  - melynek egy generációja egy bayesi modell-összehasonlítás.

# Véges állapotú automata tanítása

- Mi legyen az architektúra?
  - diszkrét optimalizálási feladat
- Adott architektúrára az élsúlyok megtalálása.
  - folytonos optimalizálási feladat
  - constrained optimization
- Keressük az optimális automatát genetikus algoritmussal  
(Csikós, 2019)
  - melynek egy generációja egy bayesi modell-összehasonlítás.
- (Borbély és Kornai, 2019)

# Hivatkozások

- Borbély, G., Kornai, A.: Sentence length. In: Proceedings of the 16th Meeting on the Mathematics of Language. pp. 114–125. Association for Computational Linguistics, Toronto, Canada (18–19 Jul 2019), <https://www.aclweb.org/anthology/W19-5710>
- Csikós, D.: Véges állapotú automaták tanulása genetikus algoritmussal. BSc Thesis, Budapest University of Technology and Economics (2019),  
<https://math.bme.hu/~csikosd/Szakdolgozat.pdf>
- MacKay, D.J.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press (2003)