esa

Post-processing of precipitation forecasts with Bernstein quantile distribution networks

John Bjørnar Bremnes
Norwegian Meteorological Institute

Probabilistic post-processing of NWP forecasts

Univariate

- What is the conditional distribution of the target variable Y given the NWP forecasts **x**?
- Distributional regression
  - Estimate Y | **x** and use it for prediction
- Multiple variables (space/time) can be estimated simultaneously

Scenarios / multivariate

Copula methods

- Ensemble copula coupling
- Observation based (Schaake shuffling)

Scenario / multivariate methods

- Conditional GANs
- Variational autoencoders
- Normalizing flows
- ….

Desired properties of (univariate) post-processing methods

- Flexible forecast distributions
  - to represent the underlying forecast uncertainty in all weather situations
- Handle many input variables/features/covariates
  - ensemble of multiple variables in space/time
  - additional variables
- Flexible relations between input variables and forecast distribution
  - multivariate function needed

Desired properties of (univariate) post-processing methods

- Flexible forecast distributions
  - to represent the underlying forecast uncertainty in all weather situations

→ **Bernstein quantile distribution**

- Handle many input variables/features/covariates
  - ensemble of multiple variables in space/time
  - additional variables

→ **Neural networks**

- Flexible relations between input variables and forecast distribution
  - multivariate function needed

→ **Neural networks**

How to model precipitation / dealing with the zeros?

Discretization / Quantization

- Probability for each category / bin
  - Training with cross-entropy loss / multinoulli likelihood
- How to discretize? Balancing of data?

Discrete model + continuous model

- Model for probability of precipitation
- Model for precipitation amounts given occurrence of precipitation
  - Training on precipitation cases only
- Combine the two using laws of probability for prediction

Censored continuous approach

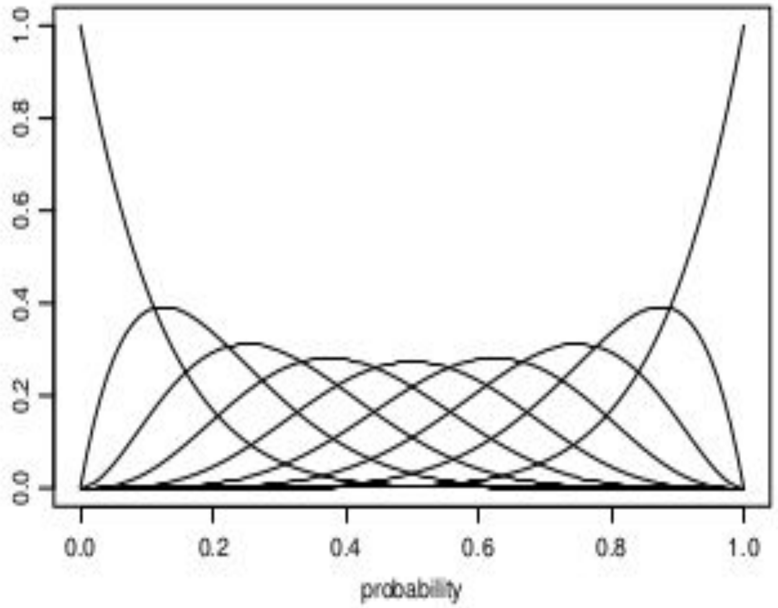
- Motivation from survival analysis
- Treat the zeros as censored values
- Introduce a latent precipitation variable that can take negative values
- Continuous model, but loss function needs to account for zeros

Definition of Bernstein quantile function distribution

$$Q(\tau|x) = \sum_{j=0}^{d} \alpha_j(x) B_{jd}(\tau)$$

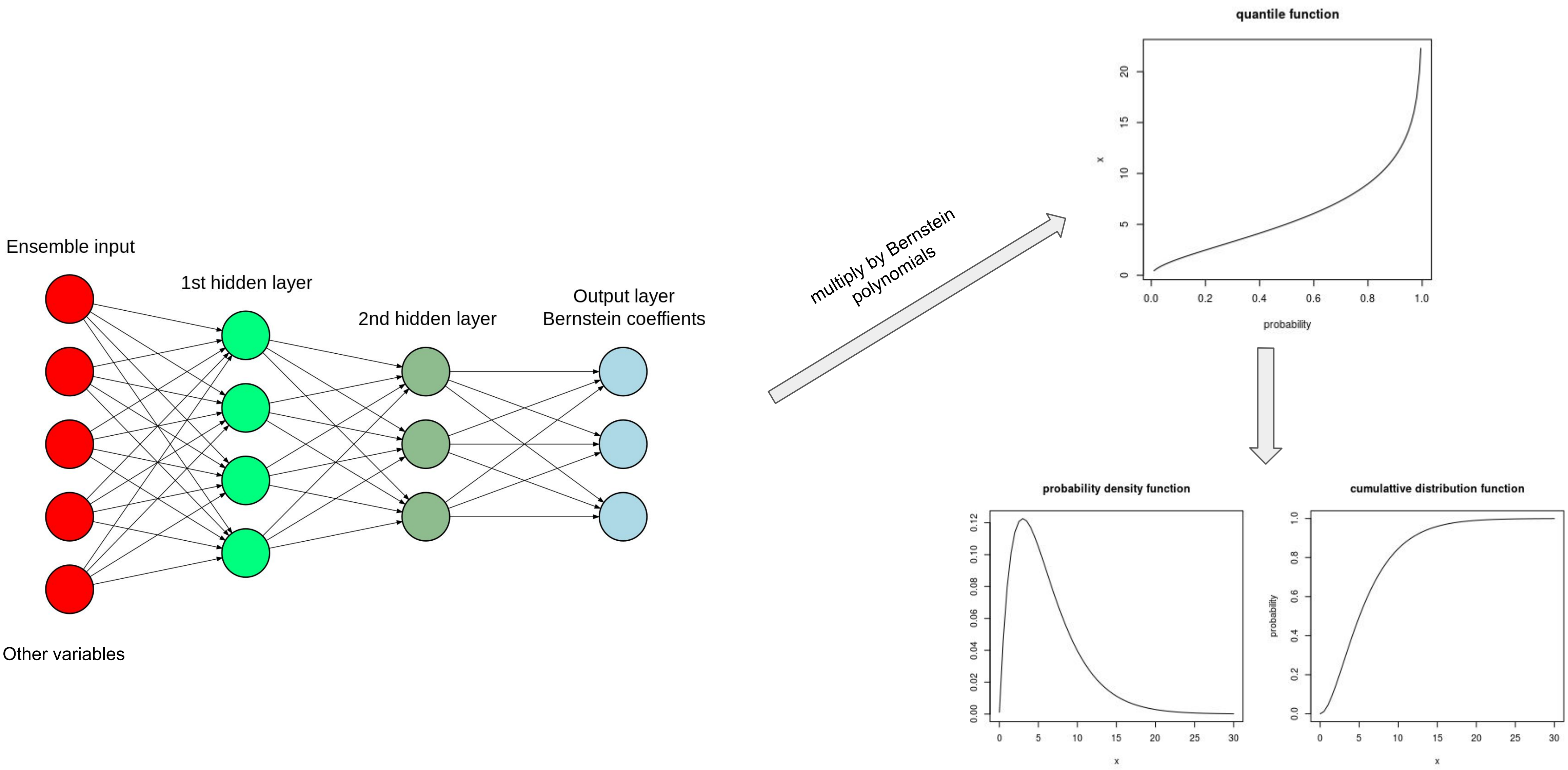
coefficients   covariates (ensemble etc.)   quantile level

Bernstein basis polynomials of degree d=8



quantile level

Note: The coefficients are functions of the input variables, here neural networks


Ensemble input
1st hidden layer
2nd hidden layer
Output layer
Bernstein coeffients
Other variables
multiply by Bernstein polynomials
quantile function
x
probability
probability density function
x
cumulattive distribution function
probability
x

How to ensure non-decreasing quantile functions?

- If the Bernstein coefficients are non-decreasing, then the quantile function is also non-decreasing. Introduce a reparameterization

$$\begin{aligned}\beta_0 &= \alpha_0\\ \beta_i &= \alpha_i - \alpha_{i-1} \geq 0, \quad i \in \{1, ..., d\}\end{aligned}$$

where the βs are the output from the neural net, or ...

- just swap quantiles (or coefficients) if necessary

Training / parameter estimation

Minimize quantile loss over a set of quantile levels $\tau_1, \tau_2, \ldots, \tau_T$

quantile loss function

training cases   quantile levels   observation   quantile function   covariates

$$\min \sum_{i=1}^{n} \sum_{t=1}^{T} \rho_{\tau_t}\left(y_i - Q(\tau_t \mid x_i)\right)$$

quantile level

$$\rho_\tau(u) = \begin{cases} u(\tau - 1) & u < 0 \\ u\tau & u \geq 0 \end{cases}$$

Note! Replacing sum over levels by integration → CRPS optimization

Censored linear quantile regression for a single quantile (τ)

Chernozhukov and Hong (2002); Friederichs and Hense (2007)

1. Create model for predicting probability of precipitation (e.g. logistic regression)
2. Create new training data set for cases where predicted **pop > 1-τ**
3. Fit quantile regression model
4. Create new training data set for cases where **fitted quantile > 0**
5. Fit quantile regression model
6. Repeat step 4 and 5 (optional)

Adaptation to BQN → censored BQN

Challenges in generalization to multiple quantiles

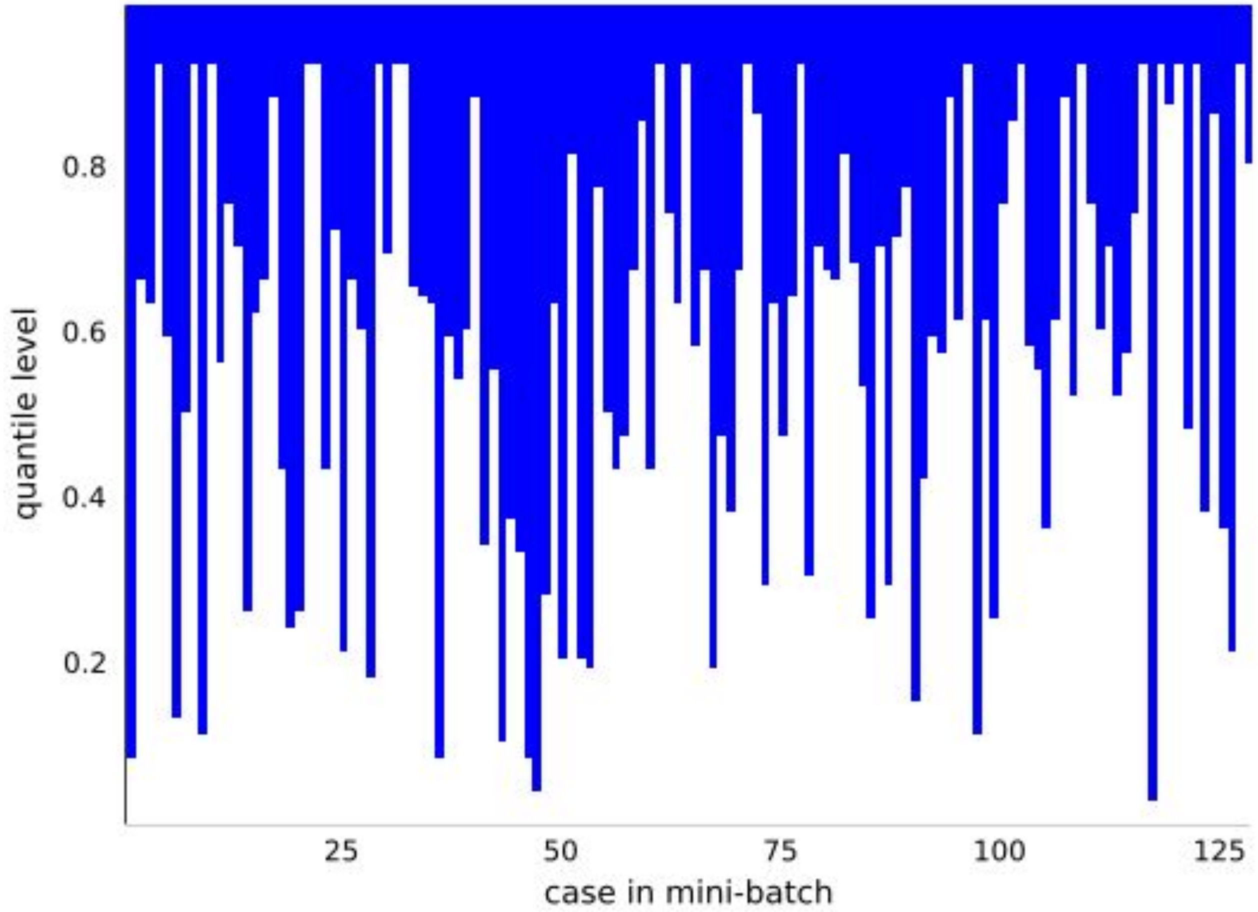
- Data points can in general no longer be left out of the training set in advance
  - For each data point the loss must be computed over a varying number of quantile levels
- Neural network training (epochs, random mini-batches)

Censored BQN approach

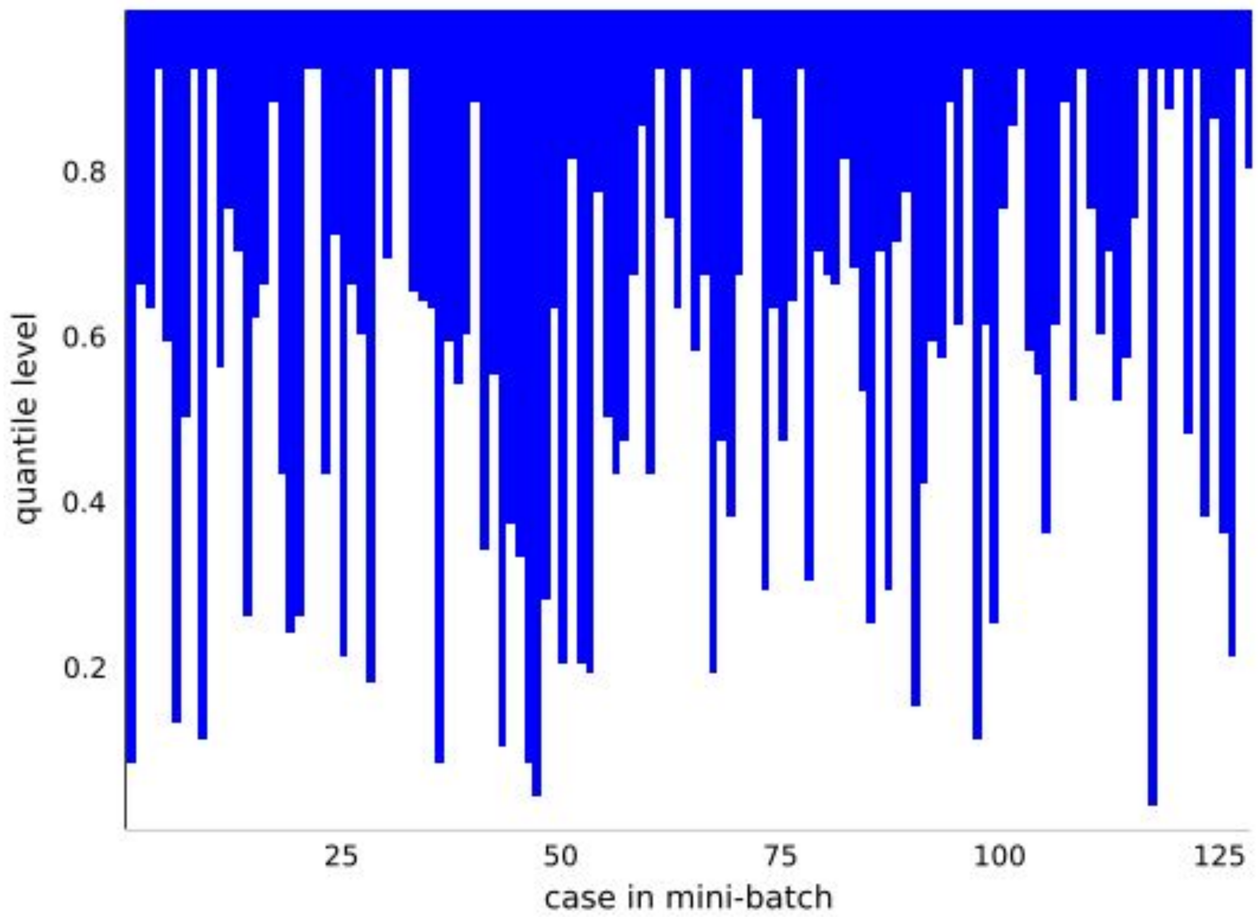
- Make a model for predicting prob. of precipitation by either
  - Creating NN model with cross-entropy loss, or
  - Using the ensemble prob. of precipitation
- 1st epoch of BQN training
  - Compute quantile loss only over levels (τ) and cases (x) where **pop(x) > 1 - τ**, $\forall x, \tau$
- Remaining epochs
  - Compute quantile loss only over levels (τ) and cases (x) where **$Q(\tau|x) > 0$**, $\forall x, \tau$

Note: $Q(\tau|x)$ is allowed to be negative, but truncated at zero for prediction

Example of which quantiles are included (blue) in computing the total quantile loss in a mini-batch



Example of which quantiles are included (blue) in computing the total quantile loss in a mini-batch



Implementation in BQN

For each mini-batch

- Create a binary matrix using
  - Probability of precip (1st epoch), or
  - Estimate of quantile function at previous iteration
- Multiply the quantile loss by the binary matrix

Are there any computational alternatives?

- Majority of elements in the loss have zero weights

Example: Synthetic precipitation data

Data from truncated shifted Gamma distribution

- Observation and 11 ensemble members generated by
  - Gamma(mean = M, varians = M*U) - s
    - M ~ Gamma(mean = 5, var = 5)
    - U ~ Uniform(0.75, 1.25)
    - s is chosen such that desired fractions of zeros is obtained (60%, 80%, 90%)
    - Truncation at zero
  - Ensemble members → 3 covariates
    - Probability of precipitation
    - Ensemble mean
    - Ensemble standard deviation
- Data sets
  - 50,000 cases for training
  - 50,000 cases for model validation/selection
  - 100,000 cases for testing

Models

- NN for pop + BQN for precipitation amounts on precipitation cases
- BQN without censoring to account for zeros
  - predictions truncated at zero
- Censored BQN with NN pop (1st epoch)
- Censored BQN with ensemble pop (1st epoch)

Method details

- Bernstein polynomials of degree 12
- Neural networks with one hidden layer (32 nodes)
  - cross-entropy loss for pop.
  - (censored) quantile loss averaged over levels 0.01, 0.02, …, 0.99 for BQN
  - batchsize 128, initial learning rate 0.001 (ADAM)
  - best model over 100 epochs chosen

60% zero precipitation cases

	Quantile Skill Score (%)					Brier Skill Score (%)	
	overall	5%	95%	Ens. median equal 0.0	Ens. median upper 1%	overall	Ens. median equal 0.0
PoP + Amount models	-4.82	-1.12	-10.02	**-4.37**	**-5.15**	-6.61	-6.92
No censoring	-4.96	-1.51	-9.98	-4.70	-5.53	-56.14	-75.64
Censoring w/ens. pop	**-4.80**	-1.02	**-9.80**	**-4.37**	-5.33	**-6.57**	**-6.88**
Censoring w/pop model	-4.84	**-1.01**	-9.89	-4.39	-5.38	-6.62	-6.95

Data-generating distribution as reference forecast
High scores best

80% zero precipitation cases

	Quantile Skill Score (%)					Brier Skill Score (%)	
	overall	5%	95%	Ens. median equal 0.0	Ens. median upper 1%	overall	Ens. median equal 0.0
PoP + Amount models	**-4.82**	-0.40	-14.45	**-4.59**	**-5.78**	**-6.92**	**-7.05**
No censoring	-5.22	-0.81	-15.41	-5.12	**-5.78**	-124.48	-152.10
Censoring w/ens. pop	-4.86	**-0.38**	-14.54	-4.61	-5.97	-6.97	-7.06
Censoring w/pop model	**-4.82**	-0.41	**-14.39**	-4.60	-5.81	-6.97	**-7.05**

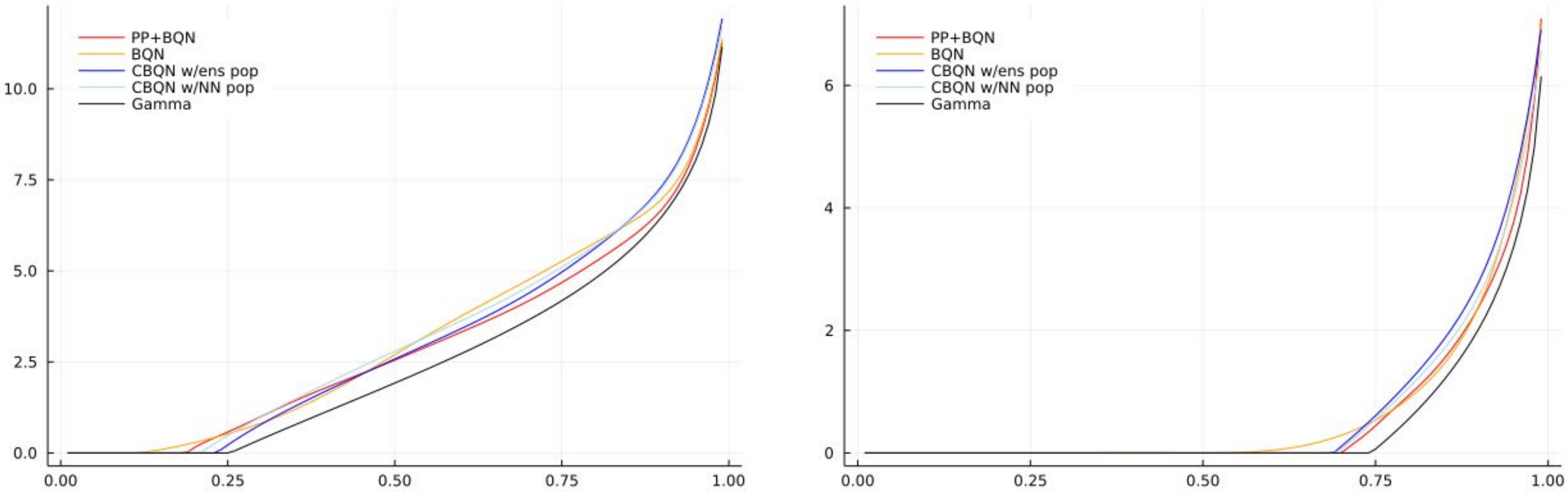
Data-generating distribution as reference forecast
High scores best

90% zero precipitation cases

	Quantile Skill Score (%)					Brier Skill Score (%)	
	overall	5%	95%	Ens. median equal 0.0	Ens. median upper 1%	overall	Ens. median equal 0.0
PoP + Amount models	-4.70	-0.35	-16.75	**-4.52**	**-6.24**	-7.09	-7.14
No censoring	-5.24	-2.95	-16.51	-5.17	-6.49	-158.65	-175.68
Censoring w/ens. pop	-4.70	-0.35	**-16.41**	**-4.52**	-6.59	-6.85	**-6.88**
Censoring w/pop model	**-4.66**	**-0.31**	-16.57	-4.53	-6.46	**-6.83**	**-6.88**

Data-generating distribution as reference forecast
High scores best

Forecast examples



Example: 6h-precipitation forecasting with ECMWF ENS reforecasts

Data

- At 48 Norwegian stations
- 00+42, +90, +138h 11-member forecasts
- 2010 - 2019 (2018 for validation and 2019 for testing)
- Input variables
  - Site id (discrete embedded)
  - Ensemble means of total precipitation, convective precipitation, total column cloud liquid water, CAPE, wind speed 700 hPa
  - Standard deviation of total precipitation
  - Probability of precipitation (total)

Models

- NN for pop + BQN for precipitation amounts on precipitation cases
- BQN without censoring to account for zeros
  - predictions truncated at zero
- Censored BQN with NN pop (1st epoch)
- Censored BQN with ensemble pop (1st epoch)

Method details

- Bernstein polynomials of degree 12
- Neural networks with one hidden layer (32 nodes)
  - cross-entropy loss for pop.
  - (censored) quantile loss averaged over levels 1/12, 2/12, …, 11/12 for BQN
  - embedding of size 6 for site id
  - batchsize 128, initial learning rate 0.001 (ADAM)
  - best model over 250 epochs chosen

00+42h

	Quantile Score					Brier Score	
	overall	5%	95%	ENS median equal 0.0	ENS median upper 1%	overall	ENS median equal 0.0
PoP + Amount models	**0.2215**			0.0097	2.0381	**0.1441**	0.0767
No censoring	0.2218			0.0099	2.0386	0.6092	0.9158
Censoring w/ens. pop	0.2233			**0.0093**	2.0186	0.1479	**0.0736**
Censoring w/pop model	0.2224			**0.0093**	**2.0170**	0.1476	0.0743
ECMWF ENS ref.	0.2907			0.0095	2.5040	0.3758	0.1008

Lower scores best

00+90h

	Quantile Score					Brier Score	
	overall	5%	95%	ENS median equal 0.0	ENS median upper 1%	overall	ENS median equal 0.0
PoP + Amount models	0.2577			0.0133	1.4959	0.1861	0.0835
No censoring	0.2579			0.0136	1.5407	0.5721	0.7804
Censoring w/ens. pop	0.2598			**0.0130**	1.5022	0.1831	**0.0796**
Censoring w/pop model	**0.2576**			**0.0130**	**1.4720**	**0.1819**	0.0798
ECMWF ENS ref.	0.3039			0.0146	1.7966	0.3697	0.1037

Lower scores best

00+138h

	Quantile Score					Brier Score	
	overall	5%	95%	ENS median equal 0.0	ENS median upper 1%	overall	ENS median equal 0.0
PoP + Amount models	**0.3579**			0.0658	**2.7449**	0.2076	0.1157
No censoring	0.3597			0.0666	2.7751	0.5839	0.7864
Censoring w/ens. pop	0.3607			**0.0656**	2.7566	**0.2007**	**0.1136**
Censoring w/pop model	0.3612			0.0657	2.7753	0.2023	0.1158
ECMWF ENS ref.	0.4061			0.0711	3.1505	0.3544	0.1372

Lower scores best

Concluding remarks

Conclusions (preliminary)

- Post-processing models need to handle the zeros in a proper way
- Similar verification results/scores for discrete+continuous approach and censored BQN (as expected)
- Censored BQN
  - Separate model for probability of precip. seems not necessary, ensemble can be used
  - Implementation by a slight adjustment of the BQN training loop

Current and future work

- Apply to gridded precipitation data (MAELSTROM project)
  - focus on computational issues on very large datasets (~10 TB)
- Compare with discretized approaches (cross-entropy models)
  - in particular for extreme weather
- How does relative forecast skill depend on the size of training data?
  - Variations between methods?

SLIDE TITLE

Test