Scribe Questions

Scribe Question 1: Project System and Objective

What is the probabilistic problem being addressed in your project? Clearly state the system objective and identify the primary sources of uncertainty involved. w

Ans:

The probabilistic problem being addressed in our project is that rainfall forecasts generate deterministic Numerical Weather Prediction (NWP) outputs, rather than showing a whole range of uncertainty. These deterministic forecasts provide a single predicted rainfall value and fail to represent uncertainty. This makes the system not very reliable to depend on for sensitive decision-making.

The system objective is to estimate the probability distribution of actual rainfall using model forecasts. This is achieved by processing raw forecast data using a Bayesian Joint Probability (BJP) framework, and producing an explicit representation of uncertainty.

The primary sources of uncertainty involved include:

- **Measurement noise**: There might be some error in data due to sensor limitation, or error in retrieval algorithm.
- **Atmospheric variability**: The dynamic nature of weather systems, especially during the Indian summer monsoon, shows unpredictability in rainfall patterns across space and time.
- **Model and parameter uncertainty**: NWP models run on simplified representations of physical processes and parameterizations, which can lead to reduced forecast skill.

Scribe Question 2: Key Random Variables and Uncertainty Modeling

Identify the key random variables in your project and describe how uncertainty is modeled for each. Clearly state any probabilistic assumptions currently being made.

Ans:

The key random variables in the project are:

1. **Forecast rainfall (X)**
   Forecast rainfall produced by the Numerical Weather Prediction (NWP) model is treated as a random variable, as its value represents an uncertain distribution.

2. **Observed rainfall (Y)**
   Observed rainfall is modeled as a random variable representing the actual rainfall outcome. As multiple observed rainfall values are possible, the observed rainfall is also considered as a random variable.

3. **Rainfall exceeding threshold event (Z)**
   Events such as rainfall exceeding a fixed threshold are modeled as a Bernoulli random variable,
   where Z=1 if Y>threshold,
   Z=0 otherwise.

- Uncertainty is modeled by assuming a joint probability distribution between the forecast rainfall and observed rainfall.
- The probabilistic relationship is expressed through the conditional distribution:
  $P(Y|X)$
  This captures the uncertainty in observed rainfall given a forecast.

Assumptions:

- As rainfall outcomes are stochastic, they can be modeled as random variables.
- The transformed forecast and observed rainfall variables are jointly normally distributed.

Scribe Question 3: Probabilistic Reasoning and Dependencies

Explain how probabilistic relationships (e.g., dependence, independence, conditionality) are used within your project system to support reasoning, inference, or decision-making.

Ans:

Dependence
Forecast rainfall **X** and observed rainfall **Y** are treated as dependent random variables, as one's probability is dependent on another. The observed rainfall outcome depends on the forecasted rainfall value, meaning that changes in **X** influence the distribution of **Y**. This dependence is modeled by assuming a joint probability distribution between **X** and **Y**.

Independence
The forecast rainfall is generated by the weather model on its own, without being directly influenced by other random variables . Therefore, it is treated as an independent input, as the model produces it first, and then we compare it with observations to build the probabilistic relationship.

Conditionality
The conditional distribution used here is P(Y∣X),
which represents the probability distribution of observed rainfall given a forecast. This conditional distribution is used to generate probabilistic rainfall forecasts and to estimate event probabilities, such as rainfall exceeding a specified threshold. Decision-making is therefore based on conditional probabilities rather than deterministic values.

Scribe Question 4: Model–Implementation Alignment

Describe how the current probabilistic model aligns with the project's implementation or experimental setup. Highlight any assumptions that influence design or evaluation choices.

Ans: The probabilistic model applied in the project is well suited to the implementation and the experimental setup. The basic model is a Bayesian Joint Probability (BJP) model that assumes that there is a joint probability distribution of the forecast and observed rainfall (**X** and **Y** respectively).

Deterministic rainfall predictions by the Numerical Weather Prediction (NWP) model are post-processed statistically with the BJP model in implementation.Rainfall data are non-normally distributed and highly skewed, and, therefore, both forecast and observed rainfall data are transformed (log-sinh transformation) to approximate normality.This directly supports the model assumption of joint normality in the transformed space.

The joint distribution parameters (means, variances, and correlation) are estimated with the help of the estimation of historical forecast-observation pairs, maximum likelihood estimation. The resultant conditional distribution **P (Y|X)** is then sampled to form probabilistic forecasts that are consistent with the system goal of characterizing forecast uncertainty.

These are some of the assumptions impacting design and evaluation options:

- The uncertainty of rainfall can have a probabilistic and not a deterministic manner of representation.

- The dependence of forecast-observation is constant during calibration.

- Behavior of seasonal rainfall is important and this results in annual (ACQ) and seasonal (SCQ) model runs.

- A leave-one-month-out cross-validation strategy, because of the short data length, is employed in validation.

The evaluation metrics (Bias percentage, CRPS, ROC, and the analysis of spread-skill) can be evaluated considering the probabilistic character of the model, so there is no disparity between theory and implementation.

Scribe Question 5: Cross-Milestone Consistency and Change

Describe the current state of your project's probabilistic model. Identify the assumptions, components, or relationships that are presently well-defined, and indicate which aspects are expected to evolve or require refinement in subsequent milestones. Briefly justify your reasoning.

Ans: At this point, there is a well-defined operational probabilistic model of the project. The next elements are clearly defined:

- Determination of important random variables: forecast rainfall (**X**) and observed rainfall (**Y**).
- Bayesian Joint Probability Model is used to indicate the dependence between **X** and **Y**.
- Assumption of joint normality following transformation.
- Conditional probability **P(Y|X)** used to create ensemble rainfall forecasts.
- Probabilistic verification metrics of performance evaluation.

But there are certain areas which are likely to change in later milestones:

- Data length and diversity: The more forecast-observation data are available, the better parameter estimation can be done and uncertainty diminished.
- Refinement of seasonal models: The seasonal calibration procedure (SCQ) can be further refined by the addition of month-by-month-calibration or regime-calibration.
- Spatial resolution: further finer grid-scale or station level applications of the model can be made.
- Extreme rainfall modeling: Existing inefficiencies in high rainfall modeling can possibly be addressed through better tail modeling.

These modifications are not surprising since probabilistic models get better as more data is provided, more knowledge is gained on the sources of uncertainty and refinement of assumptions.

Scribe Question 6: Open Issues and Responsibility Attribution

Identify any unresolved probabilistic questions or ambiguities in the project at this stage, and indicate which role(s) or tasks are responsible for addressing them in the next milestone.

Ans : At this point, a number of probabilistic problems and uncertainties remain unresolved:

1. **Extreme rainfall events representation** :
   The model is less skillful when the rainfall intensities are very high and this implies that the behavior of the tail of the probability distribution is uncertain.

2. **Effects of dearth of historical data** :
   Short data records can lead to parameter stability and robustness of the estimated joint distribution.

3. **Joint normality assumption** :
   Despite the fact that they become more normal with transformation, there might be deviations, particularly when there are extreme monsoon occurrences.

4. **Temporal non-stationarity :**
   The assumption that forecast-observation relationships are always constant over time may be broken by climate variability and varying patterns of monsoons.

Responsibility Attribution:

- **Modeling team** : Distribute assumptions of models, extreme value modeling and refine the probabilistic structure.

- **Data team**: Extend, enhance and add data sets.

- **Assessment and validation team** : Re-test verification measures and the strength of test model in alternate climatic conditions.

These will be considered in the second milestone in order to increase the predictability of forecasts and the utility of decision-making.