Analysis of Probability Determination in Bayesian Networks for Cardiovascular Risk Prediction

This document summarizes the methodologies employed to determine the probabilities for the variables in the Bayesian Networks (BNs) presented in the three cited research papers. The determination of these probabilities, often represented as Conditional Probability Tables (CPTs), is a critical step in the construction and application of any BN model.

Source	Bayesian Network Model Type	Structure Learning Method	Probability (Parameter) Learning Method	Key Distinguishing Feature
Suo et al. (2024) [1]	Weighted Survival Bayesian Network (WSBN)	Tabu Search Algorithm	Weighted Maximum Likelihood Estimation (MLE) using Inverse Probability of Censoring Weighted (IPCW) data.	Integration of a survival analysis technique (IPCW) to handle censored data.
Ordovas et. al (2023) [2]	General Bayesian Network	Fixed (Suggested Probabilistic Model)	Standard Multinomial-Dirichlet Models with Uniform Priors (Bayesian Estimation).	Combination of a large dataset with expert-defined structure and Bayesian priors.
Kong et al. (2024) [3]	General Bayesian Network	Tabu Search Algorithm	Maximum Likelihood Estimation (MLE).	Direct application of MLE on complete data, following a structure learned from the data.

Detailed Methodology by Source

1. Suo, X. et al. (2024) [1]

The study by Suo et al. developed a **Weighted Survival Bayesian Network (WSBN)** model to predict Coronary Heart Disease (CHD) risk from Electronic Health Records (EHRs), specifically addressing the challenges of **missing and censored data** common in such records.

The determination of the conditional probabilities was a multi-step process that integrated survival analysis techniques with traditional BN parameter learning:

1. **Structure Learning:** The network structure (the qualitative component) was determined using the **Tabu search algorithm** [1].
2. **Parameter Learning (Predictor Nodes):** The Conditional Probability Tables (CPTs) for the predictor variables were calculated using a modified form of Maximum Likelihood Estimation (MLE) on a **weighted data set** [1]. The weights were assigned using the **Inverse Probability of Censoring Weighted (IPCW)** method [1].

   > "Unlike the traditional parameter learning approach of directly computing the frequency distribution to estimate the conditional probability table, our method allowed us to calculate the conditional probability tables using the **weighted data set processed by the IPCW method**." [1] The IPCW method assigns weights to each sample based on the inverse probability of being censored, thereby adjusting for bias caused by censored data (e.g., loss to follow-up) [1].

3. **Parameter Learning (Outcome Node):** The final CHD outcome node's conditional probabilities were determined by **fusing the BN with a multivariate Cox Proportional Hazards (CPH) model** [1]. The predicted outcomes of the CPH model were incorporated into the BN as conditional probabilities between the CHD node and its predictor nodes [1].

In summary, the probabilities were determined by a **weighted frequency count** (a form of MLE) on the training data, where the weights were derived from a survival analysis technique (IPCW) to correct for data censoring.

2. Ordovas, J. M. et al. (2023) [2]

Ordovas et al. proposed a BN model for cardiovascular risk prediction, utilizing a combination of a large population database and expert judgment.

The determination of the probabilities (parameter learning) was achieved through a **Bayesian estimation** approach:

1. **Structure Learning:** The network structure was **fixed** based on a "suggested probabilistic model" which combined a large dataset with **expert information** [2].
2. **Parameter Learning:** The probability tables were estimated using **standard multinomial-Dirichlet models with uniform priors** [2].

> "Both the structure and the probability tables in the underlying model are built using a large dataset collected from annual work health assessments as well as **expert information**, with uncertainty characterized through posterior distributions." [2] The multinomial-Dirichlet model is a common technique for Bayesian parameter estimation in discrete BNs. The use of **uniform priors** indicates that the researchers did not impose strong initial beliefs on the probabilities, allowing the **large dataset** to primarily drive the final probability values (the expected values of the posterior distributions) [2].

3. Kong, D. et al. (2024) [3]

The study by Kong et al. focused on analyzing factors influencing Type 2 Diabetes (T2DM), Coronary Heart Disease (CHD), and their comorbidities.

The probability determination method was a straightforward application of a standard statistical technique:

1. **Structure Learning:** The BN structure was learned from the data using the **Tabu search algorithm** [3].
2. **Parameter Learning:** The parameter estimation (determining the conditional probabilities) was achieved through **Maximum Likelihood Estimation (MLE)** [3].

> "The BN structure was learned using the Tabu search algorithm, with parameter estimation achieved through **maximum likelihood estimation**." [3] MLE is the most common method for parameter learning in BNs when the data is complete. It calculates the conditional probabilities directly from the **relative frequencies** observed in the training data set [3]. The study employed a case-control design and used a dataset of 3,824 participants, which was treated as a complete dataset for the purpose of parameter estimation.

References

[1] Suo, X., Huang, X., Zhong, L., Luo, Q., Ding, L., & Li, H. (2024). Development and Validation of a Bayesian Network‐Based Model for Predicting Coronary Heart Disease Risk From Electronic Health Records. *Journal of the American Heart Association, 13*(10), e029400.

[2] Ordovas, J. M., Rios-Insua, D., Santos-Lozano, A., Lucia, A., Torres, A., Korre, M., & Camacho, J. M. (2023). A Bayesian network model for predicting cardiovascular risk. *Computer Methods and Programs in Biomedicine, 231*, 107405.

[3] Kong, D., Chen, R., Chen, Y., Zhao, L., Huang, R., Luo, L., ... & Wu, K. (2024). Bayesian network analysis of factors influencing type 2 diabetes, coronary heart disease, and their comorbidities. *BMC Public Health, 24*(1), 1267.