



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; SP-9(5): 233-239
Received: 12-02-2026
Accepted: 14-03-2026

Neha Sharma
Department of Horticulture,
School of Agriculture, Lovely
Professional University, Phagwara,
Punjab, India

Manzoor Ahmad
Department of Horticulture,
School of Agriculture, Lovely
Professional University, Phagwara,
Punjab, India

Sakshi Blagan
Department of Horticulture,
School of Agriculture, Lovely
Professional University, Phagwara,
Punjab, India

Sumit Raj Khare
Department of Horticulture,
School of Agriculture, Lovely
Professional University, Phagwara,
Punjab, India

Meragana Vatsalya Swaroop
Department of Horticulture,
School of Agriculture, Lovely
Professional University, Phagwara,
Punjab, India

Corresponding Author:
Manzoor Ahmad
Department of Horticulture,
School of Agriculture, Lovely
Professional University, Phagwara,
Punjab, India

Linking probability theory to sampling behaviour in the real world

Neha Sharma, Manzoor Ahmad, Sakshi Blagan, Sumit Raj Khare and Meragana Vatsalya Swaroop

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i5Sc.5556>

Abstract

In modern data-driven studies, it is necessary to connect probability theory with sampling behaviour in the real world to make an inference. Probability offers rigorous axioms and rigorous theorems the Law of Large Numbers, the Central Limit Theorem and the Bayes theorem of Kolmogorov which are the foundation of sampling design and quantification of uncertainty. In practice, however, practical sampling not infrequently is at variance with ideal assumptions (randomness, independence, complete frames) due to cost, availability, ethical considerations and human judgment. These exits result in selection, non-response, measurement, and cognitive bias which undermine the representativeness and inferential validity. This chapter combines the basis of probability with everyday probability and non-probability sampling methods (simple random, stratified, cluster, systematic, convenience, purposive, snowball) and describes their implications of probability. It observes computational remedies bootstrapping, Monte Carlo, MCMC, and Bayesian updating as adaptable methods of modeling intricate dependencies, hierarchical arrangements, missing facts and small samples. The technological and methodological trends engaging AI-optimized smart sampling, IoT sensors, adaptive and sequential designs, and crowdsourced citizen-science information are also discussed in the chapter and imply trade-offs between scale and bias. It has some practical recommendations that would enhance reliability, generalizability, and policy relevance of ecological, agricultural, public health, and social science studies through combining probabilistic modelling with design-aware data collection, risk assessment, and transparent reports.

Keywords: Probability theory, sampling behaviour, sampling bias, bayesian inference, Monte Carlo methods, bootstrapping, adaptive sampling, AI-enabled sampling, citizen science

1. Introduction

The basis of the statistical science and data-driven research is probability theory and sampling. The mathematical basis of quantifying uncertainty is given by probability theory, and systematic methods of choosing representative subsets of populations is given by sampling theory. These have been used together to establish the foundation of inference, prediction, and decision-making in various fields, including agriculture, social sciences, epidemiology, economics, and artificial intelligence. Nevertheless, as opposed to what probability theory presupposes, such as randomness, independence and perfect sampling frames, sampling behaviour in the real world often does not follow these assumptions because of practical, behavioural, and logistical constraints. Probability models in the theory of statistics present formal axioms and mathematical structures of the study of random phenomena. The axiomatic basis of Kolmogorov made probability a serious branch of mathematics and made it possible to derive such fundamental results as the Law of Large Numbers and the Central Limit Theorem, which justify statistical inference using samples (Kolmogorov, 2018) [15]. Sampling theory, in turn, transformed these probability principles into operational methods of choosing a subset of data on a finite or infinite population. The basic assumption is that, the selected sample should be fundamental enough to give dependable estimates of the population parameters, as long as probability-based sampling mechanisms are used. In spite of such theoretical background, in practice sampling usually takes place in imperfect situations. Cost, time, accessibility, and ethical constraints often require a researcher to use non-probability methods of sampling,

resulting in biases and lack of generalizability of the results (Shamsudin *et al.*, 2024) ^[24]. In addition, systematic errors, including selection bias and non-response bias, may be introduced in human decision-making and institutional practices and do not assume probabilities. Knowledge of how probability theory relates or fails to relate to the sampling behaviour in the real world is thus of great importance to enhance validity and reproducibility in research.

The disjunction between sampling practice and probability theory in the world of modern data-intensive applications has been redoubled. With big data, machine learning, and digital sensing technologies, the amount of data is extremely large and does not necessarily assess randomly, so the question of representativeness and inferential validity becomes a concern (Zayed, & Ba-Alwi, 2025) ^[29].

Equally, spatial heterogeneity, resource limitation and participation by farmers affect field sampling and is a problem in agriculture and environmental sciences where classical sampling assumptions are challenged. This is necessary to bridge the theoretical probability models and realities of sampling to provide strong scientific inferences and research that is policy-relevant. It is therefore important that the theoretical backgrounds of probability should be translated through practical sampling behaviour to make certain that scientific research is valid and reliable. A comprehensive insight into the transformation of probabilistic assumptions into practical sampling activities can improve experimental, survey, and observation study designs, especially when the environment is complex and rich in data.

2. Fundamentals of probability theory

The mathematical model of uncertain and random phenomena in nature and society has its mathematical basis in probability theory. It provides a quantitative model of uncertainty quantification, statistical inferences, and decision-making on sampling. Most of the statistical procedures in the design of experiments, survey research, and current uses of data science

are grounded in the principles of probability (Zhang *et al.*, 2020) ^[30].

2.1 Basic Concepts

2.1.1 Random Experiments

A random experiment refers to a process or a phenomenon whose result is not predetermined, yet the number of possible results is known. Tossing a coin, measuring crop yield in a field plot or measuring disease incidence in a population are examples. Probability modelling is founded on random experiments in theoretical and practical studies.

2.1.2 Sample Space and Events

The set of all possible outcomes of a random experiment is referred to as the sample space (Ω) and an event is a subset of the sample space that corresponds to one or more outcomes of interest to a random experiment. The events are simple (they have only one result) and compound (they have more than one result), and they can be mutually exclusive or overlapping with each other based on their structure.

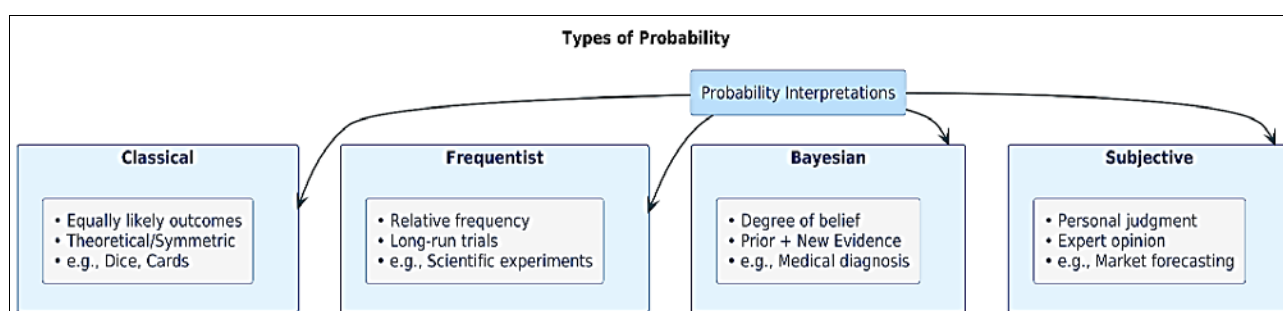
2.1.3 Axioms of Probability (Kolmogorov Framework)

The modern foundation of probability theory is based on Kolmogorov's axiomatic system, which defines probability as a function satisfying three axioms:

1. **Non-negativity:** The probability of any event is non-negative.
2. **Normalization:** The probability of the sample space is equal to one.
3. **Additivity:** For mutually exclusive events, the probability of their union is the sum of their probabilities.

These axioms provide a rigorous mathematical basis for probability and support the derivation of most probabilistic results used in statistical inference.

2.2 Types of Probability



2.2.1 Classical Probability

The classical probability is premised on equally probable events and is normally used in idealized situations like games of chance. It states that probability is the number of positive results to all the possible results when there is a sense of symmetry as well as fairness. Although it is a helpful tool in theory, classical probability cannot be applied to the complex systems of the real world.

2.2.2 Frequentist Probability

Frequentist probability views probability as the relative frequency of occurrence of an event in a series of repeated trials. Most traditional statistical inference methods are based on this interpretation, hypothesis testing and confidence intervals. It also assumes that probabilities are known yet unknown

parameters that may be estimated using repeated sampling.

2.2.3 Bayesian Probability

Bayesian probability considers probability to be an amount of belief which is revised when new information is received. It uses the previous information and the observed data by applying the Bayes theorem which is especially helpful where there is not much data or expert knowledge. Adaptive decision-making is becoming more and more popular in agriculture, epidemiology, and machine learning as a use of Bayesian methods.

2.2.4 Subjective Probability

Subjective probability is based on personal or expert opinion on the probability of occurrence of an event. It is widely applied as risk assessment, policy analysis as well as expert systems where

empirical data can be limited. Though subjectivity, it can be measured by formalizing it with Bayesian frameworks in order to facilitate rational decision-making.

2.3 Key Theorems

2.3.1 Law of Large Numbers (LLN)

The Law of Large Numbers claims that with the increase in sample size, the sample mean approaches the population mean. This theory explains the application of large samples in a study and the basis of the accuracy of statistical approximations that are created based on the sampling processes.

2.3.2 Central Limit Theorem (CLT)

According to the Central Limit Theorems, the distribution of the sample mean is similar to that of a normal distribution as the sample size is large irrespective of the population distribution. The given theorem is the core of statistical inference since it allows applying the normal-based techniques to testing hypotheses and estimating confidence intervals in various practical tasks of the real world.

2.3.3 Bayes’ Theorem: The Bayes theorem is a mathematical

property of updating probabilities using new evidence. It is the connection between prior probability, likelihood and posterior probability which is the basis of Bayesian inference. The Bayes theorem has extensive uses in predictive modelling, medical diagnostics, AI systems, and in agricultural risk assessment.

3. Sampling Theory: Concepts and Methods

A distinct separation is also made between populations and samples, as between parameters and statistics in statistical theory and practice in support of valid inference and decision-making. The population is the entire population that is going to be studied and a sample is a subset that is selected in order to estimate characteristics of the population. Population parameters are fixed and unknown parameters whereas the sample statistics are the random variables that change according to the sampling process. Also, sampling creates uncertainty in the terms of sampling error that is created by random variability, and sampling bias created by systematic flaws in the sampling design (Flinton, 2020) ^[9]. These basic ideas are critical in connecting probability theory and sampling behaviour in the real world since they define whether research results can be depended upon or not. Table 1.

Table 1: Population vs Sample, Parameters vs Statistics, and Sampling Error & Bias

Sr. No	Core Concept	Conceptual Meaning	Probabilistic Role	Applied Research Illustration	Reference
1	Population	The entire probabilistic universe in question.	Given the actual probability distribution and parameters.	The entire Indian horticultural farms; worldwide weather records.	Agarwal <i>et al.</i> , 2024 ^[1]
2	Sample	A subset sampled out of the population and probabilistic.	Foundations to estimation of unknown distributions.	300 farms were chosen; 1,000 trial plot soil samples.	Braidek <i>et al.</i> , 2007 ^[7]
3	Parameter	Real yet unknown number of population.	Target of probabilistic inference (μ, σ^2, p).	The average fruit production in the country; the rate of disease prevalence.	Jarrad <i>et al.</i> , 2015 ^[14]
4	Statistic	Estimator of a parameter obtained by use of a sample.	Sampling distribution of a random variable.	Mean yield of a sample ($\bar{x}$); p^* .	Işık <i>et al.</i> , 2005 ^[13]
5	Sampling Error	Statistic and parameter random deviation.	Measured in probability distributions (e.g., CLT).	Value census versus survey estimate.	Nirel & Glickman, 2009 ^[20]
6	Sampling Bias	A priori bias in sampling.	Breaches inferential validity and probability assumptions.	Omitting small farmers; online-alone surveys in rural research	Bradley, 2023 ^[6]

3.1 Sampling Techniques

The main types of sampling techniques can be further divided into probability sampling and non-probability sampling, based on the fact that the choice of units is based on a known probabilistic mechanism. Probability sampling techniques are based on probability theory and offer a scientific approach to statistical inference, and non-probability sampling techniques are usually motivated by practical factors and frequently applied in exploratory and qualitative studies.

3.1.1 Simple random sampling: The most basic probability sampling method is simple random sampling where all the units within a population have an equal and independent probability to be selected. This is the theoretical basis of most statistical models and is used to guarantee an unbiased estimation of population parameters. It, however, necessitates a complete and precise sampling frame which is frequently hard to come by in the practical world.

3.1.2 Stratified sampling: In stratified sampling, one uses a subdivision of the population into similar subgroups referred to as strata, and samples are taken out of each stratum. The method enhances accuracy of estimates because the sample represents significant subgroups. Stratified sampling has many applications in agricultural survey, demographic survey, and socioeconomic

studies where it is known that populations are heterogeneous.

3.1.3 Cluster sampling: Cluster sampling splits the population into clusters e.g. geographical areas or institutions and randomly chooses complete clusters to study. This method is efficient and economical when it comes to conducting surveys involving large participants, but it could enhance the sampling error in case the clusters are heterogeneous in nature. The cluster sampling is widely used in national health surveys, educational studies, and ecological research.

3.1.4 Systematic sampling: Systematic sample draws all the k th units of an ordered list following a random starting point. This is an easy technique to do and operationally effective, although it may be biased in case there exists an underlying periodic structure of the population. Systematic sampling is also common in industrial quality control and sampling in an agricultural field.

3.1.5 Non-probability sampling: Non-probability sampling techniques are not based on random sampling, and the likelihood of sample inclusion of each unit is unidentified. The convenience sampling method is used when the researcher chooses units that are readily available to him or her and is usually adopted in initial research and surveys conducted by students (Golzar *et al.*, 2022) ^[10]. Mostly, it is convenient and

cost-effective; however, it can be highly biased and less generalized.

3.1.6 Purposive (or judgmental) sampling: Purposive (or judgmental) sampling P purposive (also judgmental) sampling involves the researcher picking up the units based on his/her knowledge and predetermined variables. The method is common in qualitative research, case studies, and other assessments that involve experts, as certain information or features are necessary.

3.1.7 Snowball sampling: Snowball sampling is a network-based technique where the current members get other members through their social or business circle. The approach serves well when it comes to studying hidden or inaccessible groups, e.g. marginalized communities or rare diseases, but it is subject to network and selection biases.

4. Real-World Sampling Behaviour

Although probability theory gives a perfect picture of sampling, in reality, sampling behaviour is usually influenced by pragmatic constraints and rational choices of man and also by contextual restrictions. Sampling in applied research is not often done in purely random mechanisms, and non-probabilistic assumptions are frequent. This is important to understand these deviations and understand empirical data and enhance validity of statistical inference.

4.1 Human and Practical Constraints

Cost, time, accessibility and logistical feasibility often have an impact on the sampling choices made. Probability-based surveys and experiments deployed at large scales are generally resource-intensive in both material and human capital, which means that they cannot be practiced in less developed areas, or other resource-constrained contexts. As a result, researchers can minimise sample sizes, limit the scope of sampling, or use simplified sampling designs, and this can impact on statistical power and representativeness. Another significant limitation is accessibility especially in the agricultural, environmental and social research. Infrastructure as well as terrain and geographical demands can restrict access to some populations or research areas (Talpur *et al.*, 2013) ^[26]. The logistical problems like transportation, storage and data collection infrastructure also influence sampling strategies and can add systematic deviation of ideal random sampling frameworks.

4.1.1 Ethical and regulatory constraints: Ethic and regulatory limitations also contribute a lot towards influencing sampling behaviour. Ethical guidelines and the need to provide informed consent, as well as data protection regulations in medical and social science research, limit the choice of research participants and data collection techniques. Sampling frames can be constrained by regulatory approvals by institutional review boards or governmental agencies and may come with other limitations on the design of the study (Scheibner *et al.*, 2020) ^[23]. Although these ethical considerations are necessary to safeguard the participants, they in most cases lead to non-random or limited sampling procedures.

4.2 Behavioural Biases in Sampling

Behavioural biases: Deviant behavioural biases not based on probabilistic assumptions. The selection bias is one of the most prevalent biases whereby some units are more or less likely to be included in the sample because of some systematic reasons. As an illustration, when only urban areas are surveyed there will be

underrepresentation of other populations (rural population) thus give biased estimations.

4.2.1 Non-response bias: Non-response bias occurs when the sampled respondents fail to respond or decline to participate and their characteristics are in systematic differences with the respondents. This is especially an issue with social surveys, health studies, and online data collection where the response rates tend to be low.

4.2.2 Measurement bias: Measurement bias is a situation where the data gathered is not a representation of the actual values of a variable of interest. This may be due to ill-constructed questionnaires, the errors of the instruments or the effect of interviewers. Even random sampling can lead to a biased statistical inference by measurement bias.

4.2.3 Cognitive biases: Sampling decisions are determined by cognitive biases. Due to the biases of convenience and availability, researchers can subconsciously choose convenient or familiar sampling location, population, or data source. These cognitive aspects emphasize the human aspect of sampling behaviour and emphasize the incompleteness of theoretical probability models and their practical research application (Behimehr and Jamali, 2020) ^[4].

4.3 Sampling in Field Conditions

4.3.1 Agricultural field trials: Spatial heterogeneity, farm management practices, and environmental variability have an effect on the agricultural field trial sampling. Quasi-random or purposive sampling designs are usually caused by the selection of the plot, the involvement of farmers, and the limitation of resources (Nyarindo *et al.*, 2024) ^[21]. Randomized block design is common in field experiments in horticulture and crop science to reduce variability, but it is hardly ever attained because of practical field conditions.

4.3.2 Medical and public health surveys: The sampling is influenced by ethical considerations, the population accessibility, and the trust of the population because medical and public health surveys are not done without a purpose. Multistage cluster sampling and stratified sampling are employed in large-scale epidemiological studies to achieve a trade-off between representativeness and feasibility (Thomson *et al.*, 2020) ^[27]. Nevertheless, non-response, attrition and reporting biases are also critical issues in health studies.

4.3.3 Environmental monitoring: Spatial and temporal sampling are important in environmental monitoring as they are used to evaluate ecosystems, climate variability and biodiversity. The sampling sites are usually based on accessibility, finance, and monitoring systems hence irregular coverage. The use of remote sensing and sensor networks is becoming more popular to supplement the traditional sampling and minimize the spatial bias (Mulatu *et al.*, 2017) ^[18].

4.3.4 Social science and market research surveys: The field of social science and market research survey, sampling behaviour depends on the heterogeneity of the population, the mode of the survey (online, telephone, face-to-face, etc.), and the goodwill of the respondents. Common sampling methods include convenience and snowball sampling which is often a method of non-probability sampling (Szolnoki and Hoffmann, 2013) ^[25]. Although these techniques are informative, they do not allow a

lot of generalization and must be interpreted with caution.

5. Linking Probability Models with Real Sampling Practices

The use of probability models, in turn, gives an opportunity to conceptualize uncertainty and randomness that occurs when generating data. In practice, however, sampling in a real-life situation is seldom ideal according to these theoretical presumptions. The need to bridge probability models and practical sampling behaviour is critical to understanding empirical evidence and creating sound studies and a better way of decision-making, and across-disciplinary levels.

5.1 Ideal vs Practical Sampling

Theoretical statistics: Under the assumption of an equal probability of selection of each unit in the population, and an independent and identically distributed (i.i.d.) process of observation, ideal sampling is assumed. With these assumptions, probability models like binomial, normal and Poisson are correct models that characterize sampling behaviour and enable correct statistical inference. However, in the field, the practical sampling usually does not follow these assumptions. There might be no complete sampling frames, the population might be inaccessible, and random selection can be limited because of logistical constraints (Williams and Brown, 2019) ^[28]. Consequently, researchers often use stratified, cluster or non-probability sampling that do not assume simple random sampling. Violation randomness is the unequal or unknown selection probability and violation of independence is where sampled units are correlated in the same village (farmers), hospital (patients), etc. The result of these dependencies is the formation of clustered or hierarchical data structures, which cannot be inferred with standard models, and dangerous biases in inference are eliminated by using special probabilistic models like mixed-effects models or hierarchical Bayesian models.

5.2 Probabilistic Modelling of Sampling Errors

Sampling also comes with the problem of uncertainty that is determined using sampling distributions. A sampling distribution is a probability distribution of a statistic (sample mean or proportion, etc.) when repeated samples are used. The Central Limit Theory gives the theoretical explanation of how many sampling distributions can be found to be close to the normal distribution when the sample size is large. The sampling distributions generate confidence intervals on which there is a range of values that population parameters can be. They measure the uncertainty of sampling and are commonly applicable in agricultural research, research on the health of the population, and the social sciences to state the accuracy of estimates (Hazra, 2017) ^[11]. Probability models are applied to test hypotheses by determining the probability of observed data given a null hypothesis in hypothesis testing. The methods of t-tests, chi-square tests, and inference by regression are based on the assumptions of probabilities of sampling distributions. Non-observance of these assumptions in the actual sampling in the real world may however result in exaggerated error rates and erroneous conclusions. Bootstrap and resampling procedures have been gaining relevance to deal with the complex sampling structures and small sample sizes. These methods are computational methods that approximate the sampling distributions by resampling observed data repeatedly and do not make strong parametric assumptions (Fieberg *et al.*, 2020) ^[8]. The concept of bootstrapping is rather popular in contemporary data science, agriculture, and biomedical research to estimate uncertainty in complex models.

5.3 Bayesian Approaches in Real Sampling

Bayesian probability gives an adaptable model of combining previous understanding with sampled information. The former data could be expert opinion, past information or a previous piece of research and is formally included into probabilistic models. This method is very useful when there are limited data, sampling is expensive, and the population is inaccessible. Bayesian updating will update prior beliefs with new data as they are obtained to generate the posterior distributions, which is an expression of new uncertainty about the parameters. This learning process is similar to the real-world research because data collection and decision-making are not performed in one experiment but in the sequence (Ortoleva, 2024) ^[22]. Decision-making under uncertainty is another aspect that the Bayesian techniques facilitate through connecting probability models and utility functions as well as risk analysis and assessment. Bayesian models are applied in agriculture to predict yields and assess pests risks, in the field of public health to monitor diseases and evaluate treatment, and in the field of the environment to forecast climate and ecosystems. The Bayesian paradigm is therefore an effective compromise of theoretical probabilities and applied sampling behaviour in complicated and uncertain settings.

6. Simulation and Computational Approaches

Developments in computing capabilities and algorithms have changed the application of probability theory to sampling and inference. Simulation-related techniques enable researchers to analyze complicated probabilistic solutions, sample techniques, and quantify uncertainty in situations where an analytical solution cannot be obtained. Such methods are useful especially in the sampling in the real world, where classical probability model assumptions are frequently not met.

6.1 Monte Carlo Sampling

Monte Carlo sampling denotes a type of computational approaches, which make use of random sampling to estimate numerical outcomes. Monte Carlo methods can be used to estimate the parameters, integrals, and uncertainties of complex systems by repeatedly sampling random samples a given probability distribution. Monte Carlo simulations have been common in sampling theory to compare sampling designs, estimate properties of estimators and investigate the effects of sampling bias and variability. The methods play a crucial role in the areas of climate modelling, finance, agriculture, and physics where analytical solutions are hard or impossible to obtain (Alizadeh, 2022) ^[2].

6.2 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) techniques are monte Carlo sampling techniques that provide samples of non-trivial probability distribution by using Markov chains. The methods of MCMC, including the Metropolis Hastings and the Gibbs algorithms, are essential in Bayesian statistics, where closed-form solutions of the posteriors are typically unavailable (Malhotra, 2014) ^[16]. In real life sampling, MCMC is applied to represent hierarchical data structures, spatial dependence, as well as missing data issues. Through these, researchers are able to add uncertainty and dependence, as well as prior knowledge, to probabilistic models, and are therefore especially practical in epidemiology, ecology, and agricultural modelling.

6.3 Bootstrapping and Jackknife Methods

Non-parametric methods of estimating sampling distributions

and uncertainty are resampling methods that include bootstrapping and jackknife. Bootstrapping is a type of statistical analysis that uses repeated replacement sampling of the observed data and is used to estimate the distribution of a statistic and the jackknife is a methodical way of estimating the stability of an estimator by omitting one or more observations (Aziz *et al.*, 2025) ^[3]. These techniques are particularly useful in the case that the theoretical sampling distributions are not known or when the sample size is small. In recent studies, bootstrapping has become common in regression analysis, machine learning model testing, and the complicated survey analysis.

6.4 Role of AI and Digital Tools in Modern Sampling

The modern sampling practice is also being influenced by artificial intelligence (AI) and digital technologies. The machine learning is applied to optimize the sampling designs, identify the presence of bias in data sets, and forecast the absence of data or unobserved data. The digital tools of remote sensing, IoT sensors, mobile surveys, and big data platforms are used to collect data on a large scale, which is not restricted to the context of traditional sampling. Precision sampling in agriculture in the context of AI is used to merge sensor data with probabilistic models to enhance the estimation of yields and monitoring of pests (Aziz *et al.*, 2025) ^[3]. Digital surveys and online platforms are also quick to collect data in the social sciences and in the field of public health, but they present novel issues to do with representativeness and selection bias.

7. Future Directions and Emerging Trends

The broad and fast development of digital technologies, the use of artificial intelligence, and data-driven systems is altering the way probability theory and sampling are implemented in the field of real-life research. The new tendencies are concerned with smart sampling models, adaptive data collection plans, and a combination of probabilistic modelling with digital platforms to increase representativeness, efficiency, and decision-making in the face of uncertainty.

7.1 Smart Sampling Using AI and IoT Sensors

Smart sampling frameworks are based on artificial intelligence (AI) and Internet of Things (IoT) sensors that are used to gather information dynamically in real time. The vast amount of data are generated in high frequency by the sensors installed in farms and monitoring stations and urban areas. AI algorithms have the potential to optimize the sampling location, sampling frequency and resource allocation according to probabilistic models and predictive analytics. Intelligent sampling lowers human bias, enhances spatial and temporal coverage, as well as allows real-time decisions to be made (Ni and Wang, 2024) ^[19]. In precision agriculture, such as site-specific irrigation, fertilization and pest control, sensors are used to apply probabilistic sampling.

7.2 Adaptive and Sequential Sampling Designs

Adaptive and sequential sampling designs are a transition of the static sampling designs to the dynamic strategy of data collection. According to adaptive sampling, the future choices of sample are made based on the observed data before, and the researcher is able to target informative regions or subpopulations (Henrys *et al.*, 2024) ^[12]. Sequential sampling frames have been extensively applied to clinical trials, quality control, and ecological monitoring in order to minimize the sample size without loss of statistical power. These designs are based on probabilistic decision rules as well as stopping criteria, which provide effective alternatives to conventional fixed-sample designs.

7.3 Citizen Science and Crowdsourced Data

Citizen science and crowdsourcing sites can be used to collect data on a large scale by using the citizens in the scientific research. People can use mobile apps, electronic platforms, and social media tools to report on things to do with biodiversity, climatic events, health conditions, and social practices (Bono *et al.*, 2023) ^[5]. Although these data sources enhance the size of the sample and the extent of coverage, they are usually associated with the non-probability sampling and self-selection bias. The combination of probabilistic models and crowdsourced data is a relatively new field of research that would help to address bias and enhance the quality of data produced by citizens.

7.4 Integration of Probabilistic Modelling with Digital Agriculture and Smart Surveys

The probabilistic modelling is an emerging trend in digital agriculture and intelligent survey systems to support the decision-making processes based on data (Mishra and Mishra, 2024) ^[17]. Sampling data are used together with climate and soil and management data in probabilistic crop models, yield forecasting systems and risk assessment tools used in agriculture. Smart surveys would be applied in social and public health studies, where machine learning and adaptive questionnaires optimize the process of selecting respondents and minimizing non-response bias. Real-time inference with probabilistic modelling on a digital platform can be scaled and facilitated by complex system policymaking.

8. Conclusion

The connection between probability theory and sampling is that, they are both involved with quantifying uncertainty and also allow making statistical inference, however, in practice, sampling practice does not match theoretical assumptions because of practical, behavioural, and contextual factors. Chapter has discussed the role of probabilistic models in the design of sampling, the effects of non-randomness and lack of independence on inference and the computational and Bayesian methods as gives us useful flexible tools to model complex sampling realities. The new dynamics of smart sampling using AI, adaptive designs, citizen science, and digital data collection platforms all provide an example of how the concept of sampling is changing in the modern era of data. Combining probabilistic theory and practical sampling behaviour is necessary to increase reliability, validity, and applicability of the empirical research in all fields of interest to enable more informed scientific and policy making in more and more complex and data-rich environments.

9. References

1. Agarwal D, Devra SJ, Tiwari A. Probability and probability distribution as decision making tool in agriculture: A review. 2024.
2. Alizadeh O. Advances and challenges in climate modeling. *Climatic Change*. 2022;170(1-2). doi:10.1007/s10584-021-03298-4
3. Aziz D, Rafiq S, Saini P, Ahad I, Gonal B, Rehman SA, *et al.* Remote sensing and artificial intelligence: Revolutionizing pest management in agriculture. *Frontiers in Sustainable Food Systems*. 2025;9. doi:10.3389/fsufs.2025.1551460
4. Behimehr S, Jamali HR. Relations between cognitive biases and some concepts of information behavior. *Data and Information Management*. 2020;4(2):109-118. doi:10.2478/dim-2020-0007

5. Bono CA, Mülâyim MO, Cappiello C, Carman M, Cerquides J, Fernandez-Marquez JL, *et al.* A citizen science approach for analyzing social media with crowdsourcing. *IEEE Access*. 2023;11:15329-15347. doi:10.1109/ACCESS.2023.3243791
6. Bradley VC. Quantifying and mitigating selection bias in probability and nonprobability samples [dissertation]. Oxford: University of Oxford; 2023.
7. Braidek J, Pennock D, Yates T. Soil sampling designs. In: *Soil Sampling and Methods of Analysis*. 2nd ed. 2007. doi:10.1201/9781420005271.ch1
8. Fieberg J, Vitense K, Johnson DH. Resampling-based methods for biologists. *PeerJ*. 2020;8:e9089. doi:10.7717/peerj.9089
9. Flinton D. Sampling errors, bias, and objectivity. In: *Medical Imaging and Radiotherapy Research: Skills and Strategies*. 2020. p. 149-165. doi:10.1007/978-3-030-37944-5_10
10. Golzar J, Noor S, Tajik O. Convenience sampling. *International Journal of Education & Language Studies*. 2022;1(2):72-77. doi:10.22034/ijels.2022.162981
11. Hazra A. Using the confidence interval confidently. *Journal of Thoracic Disease*. 2017;9(10):4124-4129. doi:10.21037/jtd.2017.09.14
12. Henrys PA, Mondain-Monval T, Jarvis SG. Adaptive sampling in ecology: Key challenges and future opportunities. *Methods in Ecology and Evolution*. 2024;15(9):1483-1496. doi:10.1111/2041-210X.14393
13. Işık F, Boos DD, Li B. The distribution of genetic parameter estimates and confidence intervals from small disconnected diallels. *Theoretical and Applied Genetics*. 2005;110(7):1236-1243. doi:10.1007/s00122-005-1957-0
14. Jarrad F, Low-Choy S, Mengersen K. Biosecurity surveillance: Quantitative approaches. 2015. doi:10.1079/9781780643595.0000
15. Kolmogorov AN. *Foundations of the Theory of Probability*. 2nd English ed. Courier Dover Publications; 2018.
16. Malhotra Y. Markov Chain Monte Carlo models, Gibbs sampling, & Metropolis algorithm for high-dimensionality complex stochastic problems. *SSRN Electronic Journal*. 2014. doi:10.2139/ssrn.2553537
17. Mishra H, Mishra D. AI for data-driven decision-making in smart agriculture: From field to farm management. In: *Artificial Intelligence Techniques in Smart Agriculture*. 2024. p. 173-193. doi:10.1007/978-981-97-5878-4_11
18. Mulatu KA, Mora B, Kooistra L, Herold M. Biodiversity monitoring in changing tropical forests: A review of approaches and new opportunities. *Remote Sensing*. 2017;9(10):1059. doi:10.3390/rs9101059
19. Ni B, Wang X. Geographic research in the AI era: Applications, methods, and prospects. *Geographical Research Bulletin*. 2024;3:103-124.
20. Nirel R, Glickman H. Sample surveys and censuses. In: *Handbook of Statistics*. 2009. p. 539-565. doi:10.1016/S0169-7161(08)00021-7
21. Nyarindo W, Muger A, Hailu A, Obare GA. Do combined sustainable agricultural intensification practices improve smallholder farmers welfare? Evidence from eastern and western Kenya. *Agricultural Economics*. 2024;55(2):296-312. doi:10.1111/agec.12816
22. Ortoleva P. Alternatives to Bayesian updating. *Annual Review of Economics*. 2024;16(1):545-570. doi:10.1146/annurev-economics-100223-050352
23. Scheibner J, Ienca M, Kechagia S, Troncoso-Pastoriza JR, Raisaro JL, Hubaux JP, *et al.* Data protection and ethics requirements for multisite research with health data: A comparative examination of legislative governance frameworks and the role of data protection technologies. *Journal of Law and the Biosciences*. 2020;7(1):lsaa010.
24. Shamsudin MF, Hassim AA, Abd Manaf S. Mastering probability and non-probability methods for accurate research insights. *Journal of Postgraduate Current Business Research*. 2024;9(1):38-53.
25. Szolnoki G, Hoffmann D. Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics and Policy*. 2013;2(2):57-66. doi:10.1016/j.wep.2013.10.001
26. Talpur MAH, Napiah M, Chandio IA, Khahro SH. Research framework focusing transportation accessibility planning, computer based modeling and transportation policy outlines for remote regions of developing world. *Research Journal of Environmental and Earth Sciences*. 2013;5(1):32-40.
27. Thomson DR, Rhoda DA, Tatem AJ, Castro MC. Gridded population survey sampling: A systematic scoping review of the field and strategic research agenda. *International Journal of Health Geographics*. 2020;19(1). doi:10.1186/s12942-020-00230-4
28. Williams BK, Brown ED. Sampling and analysis frameworks for inference in ecology. *Methods in Ecology and Evolution*. 2019;10(11):1832-1842. doi:10.1111/2041-210X.13279
29. Zayed M, Ba-Alwi FM. Big data formation, reduction, and its impact on sampling: A survey. *Sana'a University Journal for Applied Sciences and Technology*. 2025;3(1):597-603.
30. Zhang J, Yin J, Wang R. Basic framework and main methods of uncertainty quantification. *Mathematical Problems in Engineering*. 2020;2020:1-18. doi:10.1155/2020/6068203