A Predictive Approach to Missing Data from the Exponential Family

Valbona Bejleri[1], Darcy Miller[2], Kay Turner[3]
[1,2,3]USDA National Agricultural Statistics Service, 1400 Independence Ave. SW, Washington, DC 20250

Abstract
Discrete outcomes are often observed among survey responses, i.e., counts distributed as Poisson, Bernoulli, multinomial, etc. When some responses are missing, the observed data provide a foundation for predicting the unobserved values or estimating some statistic for the full sample. In this paper, data are assumed to be a random sample from a discrete distribution in the exponential family with missing at random (MAR) responses, i.e., the probability of a response missing is unrelated to the value of that response, but could be related to other variables. A distribution-based technique to compute lower and upper bounds for a missing response is developed. The algorithm makes no use of the parameter estimate. Simulations are used to illustrate the technique and to assess its efficiency. Bayesian prediction bounds constructed based on gamma, Jeffreys, and uniform priors are also discussed for comparison purposes. We tested the algorithm using actual data from USDA's National Agricultural Statistics Service's Quarterly Hog Survey.

Key Words: Bounds, Discrete distribution, Exponential family, Imputation, Predictive distribution.

1. Introduction

The existing techniques on handling missing data could roughly be categorized in complete case analysis, re-weighting (e.g., calibration), and imputation (e.g. hot-deck, mean, regression, data augmentation, fully conditional specification). Often, we want to consider limiting our imputations to a range of values to meet requirements of edits that are written for the survey or staying within the support of the variable to impute. This is handled using post-imputation adjustments and minimizing objective functions or imputing within bounds/limits. In practice, the bounds or range of plausible values are determined by the agency, organization, or analyst, often through programmed edits that flag erroneous or likely erroneous values. Literature exists on how to do both, and software programs are available to implement methodologies (Coutinho et al. (2010), De Waal (2005), Pannekoek et al (2008), Fellegi and Holt (1976), Raguhnathan (2001), Winkler (2001), Tempelman (2007), Kim et al. (2014), etc.).

Prediction is used to address many practical problems related to estimating some statistic of an unobserved sample, e.g., the mean or the range of one or more unobserved values, based on the information available from an observed sample. The literature goes back as far as Weiss (1955) who developed an approximate prediction interval ("parameter-free

confidence set") by utilizing the approximation to the normal distribution and the idea of conditioning on a sufficient statistic. Later, Weiss presented a prediction problem from the decision theory point of view (Weiss, 1961). Thatcher (1964) developed prediction intervals (frequentist and Bayesian) for the number of successes that will be observed in a future binomial experiment. Related issues on prediction intervals and their predictive likelihoods are addressed in Hahn and Nelson (1973). Barndorff-Nielsen and Cox (1975) derived a predictive distribution function based on asymptotic considerations, and Lawless and Fredette (2005) estimated a predictive distribution using simulated base methods. Predictive inferences are found in the Bayesian framework in Aitchison and Sculthorpe (1965), Aitchison and Dunsmore (1975), Hall et al. (1999), etc.

In this paper, we present a distribution-based technique to compute prediction limits/bounds to use for imputation in conjunction with known support of the variable to impute. This is conducted as a complimentary (or alternative) to methods such as using extreme respondent values, respondent percentiles, previously reported data, etc. Data are assumed to be a random sample from a discrete distribution of the exponential family (i.e., Poisson, Bernoulli, multinomial, etc.) with missing at random (MAR) responses, i.e., the probability of a response missing is unrelated to the value of that response but could be related to values of other variables. We utilize the observed information and predictive inferences to generate the most extreme data values that would be acceptable for imputation, i.e. satisfy the survey edits or fit within a support of a variable of interest. After setting up the problem in Section 2, we derive the optimal (smallest) frequentist upper prediction limit for an unobserved data Y_{nr} in Section 3 based on some predetermined probability of wrong prediction, using the information from the observed data Y_r. For a fixed θ, both the observed and the data to be predicted follow a discrete distribution from the exponential family. The upper limit of Y_{nr} is derived from the joint probability distribution of the observed and the unobserved data, $p(y_r, y_{nr}|\theta) = p(y_r|\theta) * p(y_{nr}|\theta)$. It is computed using a numerical approach based on some predetermined probability α of wrong prediction. Example 1 illustrates the technique. Bayesian prediction limits are also discussed for comparison purposes in Section 4. The optimal (lowest) Bayesian upper limit for the unobserved Y_{nr} is derived from the posterior predictive distribution of the random variable Y_{nr} given the random sample $\underline{Y_r} = (y_{r1}, y_{r2}, \dots, y_{rn})$, $p(y_{nr}|y_r) = \left[\int_\theta p(y_{nr}|\theta) p\left(\underline{y_r}\middle|\theta\right) p(\theta) d\theta\right] / \left[\int_\theta p\left(\underline{y_r}\middle|\theta\right) p(\theta) d\theta\right]$, based on some predetermined maximum probability of wrong coverage α. Parameter θ is assumed to be a random variable. Concluding remarks are presented later in Section 5.

2. Problem Setup

Let the random variable Y_r describe the observed data and Y_{nr} the missing data to be imputed, $Y = (Y_r, Y_{nr})$. Both Y_r and Y_{nr} conditioned on θ are independent

$$p_{y_r}(y_r|\theta) = h(y_r)c(\theta)exp\{w(\theta)y_r\} \tag{2.1}$$

$$p_{y_{nr}}(y_{nr}|\theta_1) = h(y_{nr})c(\theta_1)exp\{w(\theta_1)y_{nr}\}, \tag{2.2}$$

where $\theta_1 = k\theta$, $k = c^{te}$. Without loss of generality, let $k = 1$ and hence $\theta_1 = \theta$. Furthermore, $h(.)$ is a nonnegative real-valued function that does not depend on θ and $c(.)$ is a nonnegative real valued function of θ that does not depend on y_r or y_{nr}.

In this paper, we present a general algorithm of how to construct a function $u(Y_r)$ ($l(Y_r)$) that takes only integer values and that will serve as an upper (lower) bound for the values of the random variable Y_{nr}. Although this function might not be unique, the algorithm shows how to derive the unique lowest upper bound $u^*(Y_r)$ (greatest lower bound $l^*(Y_r)$) of a single nonresponse from the exponential family with respect to some error probability α (β). For the frequentist approach, these bounds are derived based on the joint distribution $P(Y_r, Y_{nr}|\theta)$ of both responses and nonresponses. For the Bayesian approach, the bounds are derived based on the posterior predictive distribution $P(Y_{nr}|Y_r, \theta)$.

3. Frequentist Approach

In this section, the upper prediction limit $u^*(Y_r)$ is computed such that the frequency of making a wrong prediction will not exceed the predefined maximum error probability α

$$P\{Y_{nr} > u^*(Y_r)\} \leq \alpha. \quad 3.1$$

The joint distribution of Y_r and Y_{nr}, $p(y_r, y_{nr}|\theta) = \left[p_{Y_r}(y_r|\theta)\right]\left[p_{Y_{nr}}(y_{nr}|y_r, \theta)\right]$, is

$$p(y_r, y_{nr}|\theta) = [h(y_r)c(\theta)exp\{w(\theta)y_r\}][h(y_{nr})c(\theta)exp\{w(\theta)y_{nr}\}]. \quad 3.2$$

We write it as a product of two terms, $p(y_r, y_{nr}|\theta) = [p_\tau(\tau|\theta)]\left[p_{Y_r}(y_r|\tau, \theta)\right]$, such that one term will express the probability distribution function of the sufficient statistic $T = T(Y_r, Y_{nr})$. Note that $T = \tau$, and its distribution, $p_\tau(\tau|\theta)$, is from the exponential family (Casella and Berger 2002, p.217):

$$p(y_r, y_{nr}|\theta) = \left\{H(\tau)c_1(\theta)exp\{w(\theta)\tau\}\right\}\left\{\frac{h(y_r)h(y_{nr})}{H(\tau)}\right\}, \quad 3.3$$

where $c_1(\theta) = [c(\theta)]^2$ and $H(\tau)$ is a normalizing factor that does not depend on θ.

Function $H(\tau)$ in 3.3 is theoretically known as long as we know the representation of $h(y_r)$ and $h(y_{nr})$. We can calculate $H(\tau)$ numerically as a cumulative sum of $h(y_r)$ and $h(y_{nr})$, even in cases when space of Y_r is unbounded. First, for the observed y_r and a very small ϵ, $0 < \epsilon < 1$, choose τ_{max} such that $1 - \sum_{y_r}[h(y_r)h(\tau - y_r)/H(\tau)] < \varepsilon$. Then, we have to go out far enough in the space of Y_r to make sure that all possible terms are included in $H(\tau)$. Thus, for a realization y_{robs} of Y_r, choose $y_{rmax} = \max\{y: y_{robs} \leq y \leq \tau_{max}\}$, such that $h(y_r)h(\tau - y_r)/H(\tau) < \epsilon$ and then compute $H(\tau)$,

$$H(\tau) = \sum_{y_r \leq \tau_{max}}[h(y_r)h(\tau - y_r)]. \quad 3.4$$

An upper limit $u(Y_r)$ for the unobserved data Y_{nr} is a function of random variable Y_r. The probability of wrong prediction at the upper limit $u(Y_r)$ does not exceed α, that is,

$$P\big(Y_{nr} > u(Y_r)\big) = \sum_\tau\left[\left\{H(\tau)c_1(\theta)exp\{w(\theta)\tau\}\right\}\Delta_\tau\right] \leq \alpha, \quad 3.5$$

where $\Delta_\tau = \sum_{y_r<\tau-u^*(y_r)}[[h(y_r)h(y_{nr})]/H(\tau)]$ depends on both τ and $u(y_r)$. We can interpret and compute Δ_τ in terms of the cumulative distribution function of the random variable Y_r given $\tau = y_r + y_{nr}$.

Among all integer-valued functions $u(Y_r)$ that satisfy the latter condition for all $\tau \geq 0$, we are interested in the function $u^*(Y_r)$ that is the smallest, $u^*(Y_r) \leq u(Y_r)$ for every $Y_r = y_r$. Choose $u^*(Y_r)$, $u^*(y_r) = \max\{y: \sum_{y_r<\tau-u^*(y_r)}[[h(y_r)h(y)]/H(\tau)] > \alpha\}$. Hence, $P(Y_{nr} > u^*(Y_r)) \leq \alpha$. The greatest lower bound $l^*(Y_r)$ of a single nonresponse Y_{nr} is derived with respect to some error probability β using a similar algorithm. Conditions on Y_r and Y_{nr} will remain the same. The only difference in this case is defining the probability of wrong prediction as $P\{Y < l^*(X)\} \leq \beta$.

3.1 Illustration: The Poisson Case

Random variables Y_r and Y_{nr} conditioned on the rate parameter λ are independent Poisson. Their distributions are given by 2.1 and 2.2, where $h(y_r) = (s_1)^{y_r}/(y_r!)$, $h(y_{nr}) = (s_2)^{y_{nr}}/(y_{nr}!)$, $s_1 = 10$ is the past time interval, $\theta = \lambda s_1$, $s_2 = 5$ is the future time interval, and $\theta_1 = \lambda s_2$, and λ is unknown.

$$P\{Y_{nr} > u^*(Y_r)\} = \sum_{\tau=0}^{\infty}\left\{\frac{(\lambda s_1+\lambda s_2)^\tau e^{-(\lambda s_1+\lambda s_2)}}{\tau!}\Delta_\tau\right\} \leq \alpha, \qquad 3.6$$

where $\tau = y_r + y_{nr} \in Z^+$, $\Delta_\tau = \sum_{y_r<\tau-u^*(y_r)}[[h(y_r)h(y_{nr})]/H(\tau)]$, and $\alpha = .05$. $H(\tau)$ is computed numerically as

$$H(\tau) = \sum_{y_r\leq\tau_{max}}\left[[(s_1)^{y_r}/(y_r!)][(s_2)^{\tau-y_r}/(\tau-y_r!)]\right]. \qquad 3.7$$

The values of τ_{max} and $u^*(Y_r)$ computed using the general algorithm for the Poisson illustration are presented in Table 1. Figure 1 demonstrates the derivations of the upper limits shown in Table 1. Each row of the matrix in Figure 1 represents a cumulative distribution function of $Y_r|\tau$, and $\Delta_\tau = \sum_{y_r<\tau-u^*(y_r)}[[h(y_r)h(y_{nr})]/H(\tau)]$, where τ changes through rows and y_r changes through columns. Figure 2 demonstrates the derivations of the upper limits shown in Table 1 when the distributions of Y_r and Y_{nr} are known to be Poisson. τ changes through diagonals while y_r changes through columns. Each diagonal represents a probability mass function of $Y_r|\tau$ on the Poisson distribution. The probability of wrong coverage is calculated as a Poisson weighted sum of Δ_τ terms indicated in blue; $\tau = y_r + y_{nr} > 0$.

The exact frequentist lower prediction limit ($g(Y_r)$) for a single future observation from a distribution in the exponential family is derived using a similar algorithm. Conditions on Y_r and Y_{nr} will remain the same. The only difference is in defining the probability of wrong prediction. In this case, the probability of wrong prediction is given by $P\{Y < g(Y_r)\} \leq \beta$, and it should not exceed some error probability β.

We tested the algorithm using actual data from the quarterly hog survey. Using a small subset of quarterly hog survey responses, we predicted the upper bounds for the missing counts of total hogs and pigs used for market and at home per county in a particular stratum whenever the information on total weight was available. The prediction bounds included the true count in 90% of the cases, and missed it in only 10%. Further testing on

larger sample sizes will help better investigate the actual probability of wrong prediction and the efficiency of the algorithm.

Table 1. The computed values of τ_{max} and $u^*(Y_r)$ when the observed $Y_r = y_r$, $s_1 = 10$, and $s_2 = 5$

y_r	0	1	2	3	4	5	6	7	8	9	10	11	12	13
τ	2	4	6	8	10	12	14	15	17	19	21	22	24	26
$u^*(y_r)$	2	3	4	5	6	7	8	8	9	10	11	11	12	13

$\tau \backslash y_r$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	NA	0.000	0.000	0	0	0	0	0	0	0	0	0	0	0
1	0.333	1.000	0.000	0	0	0	0	0	0	0	0	0	0	0
2	**0.111**	0.556	1.000	0	0	0	0	0	0	0	0	0	0	0
3	0.037	0.259	0.704	1	0	0	0	0	0	0	0	0	0	0
4	0.012	**0.111**	0.407	0.802	1.000	0.000	0.000	0	0	0	0	0	0	0
5	0.004	0.045	0.210	0.539	0.868	1.000	0.000	0	0	0	0	0	0	0
6	0.001	0.018	**0.100**	0.320	0.649	0.912	1.000	0	0	0	0	0	0	0
7	0.000	0.007	0.045	0.173	0.429	0.737	0.941	1	0	0	0	0	0	0
8	0	0.003	0.020	**0.088**	0.259	0.532	0.805	0.961	1.000	0.000	0.000	0	0	0
9	0	0.001	0.008	0.042	0.145	0.350	0.623	0.857	0.974	1.000	0.000	0	0	0
10	0	0.000	0.003	0.020	**0.077**	0.213	0.441	0.701	0.896	0.983	1.000	0	0	0
11	0	0.000	0.001	0.009	0.039	0.122	0.289	0.527	0.766	0.925	0.988	1	0	0
12	0	0	0.001	0.004	0.019	**0.066**	0.178	0.368	0.607	0.819	0.946	0.992	1.000	0.000
13	0	0	0.000	0.002	0.009	0.035	0.104	0.241	0.448	0.678	0.861	0.961	0.995	1.000
14	0	0	0.000	0.001	0.004	0.017	**0.058**	0.149	0.310	0.524	0.739	0.895	0.973	0.997
15	0	0	0.000	0.000	0.002	0.009	0.031	**0.088**	0.203	0.382	0.596	0.791	0.921	0.981
16	0	0	0	0	0.001	0.004	0.016	0.050	0.127	0.263	0.453	0.661	0.834	0.941
17	0	0	0	0	0.000	0.002	0.008	0.027	**0.075**	0.172	0.326	0.522	0.719	0.870
18	0	0	0	0	0.000	0.001	0.004	0.014	0.043	0.108	0.223	0.391	0.588	0.769
19	0	0	0	0	0.000	0.000	0.002	0.007	0.024	**0.065**	0.146	0.279	0.457	0.648
20	0	0	0	0	0	0	0	0.004	0.013	0.038	0.092	0.191	0.339	0.521
21	0	0	0	0	0	0	0	0.000	0.007	0.021	**0.056**	0.125	0.240	0.399
22	0	0	0	0	0	0	0	0.000	0.000	0.012	0.033	**0.079**	0.163	0.293
23	0	0	0	0	0	0	0	0.000	0.000	0.000	0.019	0.048	0.107	0.206
24	0	0	0	0	0	0	0	0	0	0	0	0.028	**0.068**	0.140
25	0	0	0	0	0	0	0	0	0	0	0	0.000	0.042	0.092
26	0	0	0	0	0	0	0	0	0	0	0	0.000	0.000	**0.058**

Figure 1: Entries Δ_τ of this matrix are computed as $\sum_{y_r<\tau-u^*(y_r)}[[h(y_r)h(y_{nr})]/H(\tau)]$; τ changes through rows while y_r changes through columns. The τ_{max} and y_r, indicated by red colored cells, yield the upper limit function $u^*(y_r) = \tau_{max} - y_r$.

$y_{nr}\backslash y_r$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
[0,]	NA	0.667	0.444	0.296	0.198	0.132	0.088	0.059	0.039	0.026	0.017	0.012	0.008	0.005
[1,]	0.333	0.444	0.444	0.395	0.329	0.263	0.205	0.156	0.117	0.087	0.064	0.046	0.033	0.024
[2,]	**0.111**	0.222	0.296	0.329	0.329	0.307	0.273	0.234	0.195	0.159	0.127	0.100	0.078	0.060
[3,]	0.037	**0.099**	0.165	0.219	0.256	0.273	0.273	0.260	0.238	0.212	0.184	0.156	0.130	0.107
[4,]	0.012	0.041	**0.082**	0.128	0.171	0.205	0.228	0.238	0.238	0.230	0.214	0.195	0.173	0.151
[5,]	0.004	0.016	0.038	**0.068**	0.102	0.137	0.167	0.191	0.207	0.214	0.214	0.208	0.196	0.181
[6,]	0.001	0.006	0.017	0.034	**0.057**	0.083	0.111	0.138	0.161	0.179	0.190	0.196	0.196	0.191
[7,]	0.000	0.002	0.007	0.016	0.030	**0.048**	0.069	0.092	0.115	0.136	0.154	0.168	0.178	0.182
[8,]	0	0.001	0.003	0.007	0.015	0.026	**0.040**	**0.057**	0.077	0.096	0.116	0.133	0.148	0.159
[9,]	0	0.000	0.001	0.003	0.007	0.013	0.022	0.034	**0.048**	0.064	0.081	0.099	0.115	0.130
[10,]	0	0.000	0.000	0.001	0.003	0.007	0.012	0.019	0.029	**0.041**	0.054	0.069	0.084	0.100
[11,]	0	0.000	0.000	0.001	0.002	0.003	0.006	0.011	0.017	0.025	**0.035**	**0.046**	0.059	0.072
[12,]	0	0	0	0	0.001	0.002	0.003	0.006	0.009	0.014	0.021	0.029	**0.039**	0.050
[13,]	0	0	0	0	0.000	0.001	0.001	0.003	0.005	0.008	0.012	0.018	0.025	**0.034**

Figure 2: Matrix of the probability mass function of $Y_r|\tau$, where $\tau = Y_r+Y_{nr}$. The corresponding row of the red colored cell probabilities indicate the maximum τ such that $\Delta_\tau > .05$; τ changes through diagonals. Hence, the row indexes indicate the upper limit function $u^*(y_r)$. The probability of wrong coverage is calculated as a Poisson weighted sum of Δ_τ (sum of the lower part of the diagonal) indicated in blue; $\tau = y_r + y_{nr} > 0$.

4. Bayesian Approach

The construction of the predictive density function is the essence of Bayesian prediction analysis. In this section we present general steps to compute the predictive density function and Bayesian prediction limits of a discrete distribution from the exponential family. Similar to the frequentist approach, our data consist of two parts,Y $= (Y_r, Y_{nr})$, where Y_r describe the observed data and Y_{nr} the missing data to be imputed. Both Y_r and Y_{nr} conditioned on θ are independent

$$p_{y_r}(y_r|\theta) = h(y_r)c(\theta)exp\{w(\theta)y_r\} \tag{4.1}$$

$$p_{y_{nr}}(y_{nr}|\theta) = h(y_{nr})c(\theta)exp\{w(\theta)y_{nr}\}, \tag{4.2}$$

where $h(.)$ is a nonnegative real-valued function that does not depend on θ and $c(.)$ is a nonnegative real-valued function of θ that does not depend on y_r or y_{nr}. Furthermore, θ is now considered a random variable.

Our goal is to construct a function $v^*(Y_r)$ which takes only integer values and will serve as an optimal upper bound for the imputed values of Y_{nr} with respect to some error probability α, when Y_r is observed.

For θ, we considered proper priors from the conjugate class of the exponential family of distributions that belong also to the exponential family

$$p(\theta) \propto [c(\theta)]^\eta exp\{w(\theta)\rho\}. \tag{4.3}$$

The posterior distribution of θ given Y_r would also be from the exponential family

$$p\left(\theta|\underline{y_r}\right) \propto [c(\theta)]^{\eta+n} exp\{w(\theta)\}(\rho + z), \tag{4.4}$$

where $\underline{Y_r} = (y_{r1}, y_{r2}, \dots, y_{rn})$ and $z = \sum_{i=1}^{n} y_{ri}$.

The posterior predictive distribution, $P(Y_{nr}|Y_r)$, is computed as $\int_\theta \ P(Y_{nr}|\theta)P(\theta|Y_r)d\theta$

$$P\left(Y_{nr}\middle|\underline{Y_r}\right) \propto \int_\theta \ h(y_{nr})[c(\theta)]^{\eta+n+1} exp\{w(\theta)\}(y_{nr} + \rho + z)d\theta. \tag{4.5}$$

For the observed $Y_r = y_r$, we take as the smallest upper Bayesian prediction limit $v^*(y_r)$, the $1 - \alpha$ percentile of the predictive distribution. And, as the largest lower Bayesian prediction limit $l^*(y_r)$, we take one more than the β percentile of the predictive distribution.

Jeffreys and uniform non-informative priors were adopted for the parameter. The upper prediction limits based on uniform prior and lower prediction limits based on Jeffreys prior appeared to coincide with their respective exact frequentist upper and lower prediction limits.

5. Concluding Remarks

We introduced a distribution-based (frequentist) technique to compute lower and upper bounds for a missing response from the exponential family. The algorithm makes no use of the parameter estimate. Adding these bounds as a new constraint would guarantee, with respect to some predefined error probability, that the imputed values belong to the same population as the observed values of the variable under consideration. We tested the algorithm using a small sample from the quarterly hog survey and plan to further investigate the probability of wrong prediction and the efficiency of the algorithm by testing it on larger size datasets.

Upper and lower Bayesian prediction limits were discussed. A proper prior distribution from the exponential family of distributions was adopted for the parameter. We also considered uniform and Jeffreys noninformative priors and investigated the relationship of the Bayesian prediction limits derived based on these noninformative priors with their respective frequentist prediction limits. The upper prediction limits based on uniform prior and lower prediction limits based on Jeffreys prior appeared to coincide with their respective exact frequentist upper and lower prediction limits.

References

Aitchison, J. and Dunsmore, I. R. (1975), *Statistical Prediction Analysis*, Cambridge, England: Cambridge *University Press*, p 68, 227-231.

Aitchison, J. and Sculthorpe, D. (1965), "Some Problems of Statistical Prediction", *Biometrica*, 52, 469-483.

Bejleri, V. (2005). *Bayesian Prediction Intervals for the Poisson Model, Noninformative Priors*, Ph.D. diss., American University, Washington, DC.

Barndorff-Nielsen, O. E. and Cox, David R. (1996), Prediction and Asymptotics, *Bernoulli*, Vol. 2, No. 4, pp. 319-340.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, (Second Edition), Wadsworth Group, p 66-67.

Cox, D. R. (1975), Prediction Intervals and Empirical Bayes Confidence Intervals, *Perspectives in Probability and Statistics*, ed. J. Gani, London: Academic Press, pp 47-55.

Coutinho, W., de Waal, T., and Shlomo, N. (2010), Calibrated Hot Deck Imputation Subject to edit Restrictions. Discussion paper 201016, Statistics Netherlands.

de Waal, T. (2005), Automatic Error Localization for Categorical, Continuous and Integer Data. Statistics and Operations Research Transactions 29, pp. 57-99.

Fellegi, I.P. and Holt, D. (1976), "A Systematic Approach to Automatic Edit and Imputation," Journal of the American Statistical Association, 71, 17-35.

Gelman, A., Carlin, B. J., Stern, S. H. and Rubin, B. D. (2004), *Bayesian Data Analysis*, (Second Edition), Chapman and Hall/CRC, p 39-43.

Hahn, G. J.and Nelson, W. (1973), "A Survey of Prediction Intervals and Their Applications", *Journal of Quality Technology*, Volume 5, Number 4, 178-188.

Hall, P., Peng, L. and Tajvidi, N. (1999), On Prediction Intervals Based on Likelihood or Bootstrap Methods, *Biometrika*, 86, pp 871-880.

Kass, R. and Wasserman, L. (1996), The Selection of Prior Distributions by Formal Rules, *Journal of American Statistical Association*, Vol.91, No.435, p 1343-1370.

Kim, H., et al. (2014). Multiple Imputation of Missing or Faulty Values Under Linear Constraints, *Journal of Business & Economic Statistics*, Vol. 32, No. 3, pp 375-386.

Kvam, P. H. and Miller, J. G. (2002). Discrete Predictive Analysis in Probabilistic Safety Assessment, *Journal of Quality Technology*, Vol.34, No.1.

Lawless, J. F. and Fredette, M.(2005). Frequentist Prediction Intervals and Predictive Distributions *Biometrika*, 92: 529-542.

Pannekoek, J. et al (2008). "Calibrated Imputation of Numerical Data under Linear Edit Restrictions". Working Paper No. 23, UNECE Work Session on Statistical Data Editing.

Raghunathan, T.E., Lepkowski, J.M., Hoewyk, J.V. and Solenberger, P. (2001). "A Multivariate Technique for Multiply Imputing Missing Values using a Sequence of Regression Models". Survey Methodology, 27, 85-95.

Thatcher, A.R. (1964). Relationship Between Bayesian and Confidence Limits for Prediction, *Journal of Royal Statistical Society*, Ser. B, 26, pp 176-192.

Tempelman, D. C. G. (2007), Imputation of Restricted Data. Ph.D. thesis, university of Groningen.

Weiss, L. (1955) A note on Confidence Sets for Random Variables. *Annals of Mathematical Statistics* 26: 142-144.

Weiss, L. (1961). *Statistical Decision Theory*, New York: McGraw-Hill.

Winkler, Robert L. (1972). *An Introduction to Bayesian Inference and Decision*, Holt, Rinehart and Winston, Inc., p 91.

Winkler, W. and Gor-Chung, C. (2001). Extending the Fellegi-Holt Model of Statistical Data Editing, Proceedings of the Annual Meeting of the American Statistical Association.
https://www.amstat.org/sections/srms/Proceedings/y2001/Proceed/00246.pdf