ELSEVIER

Statistics & Probability Letters 41 (1999) 247-254

STATISTICS & PROBABILITY LETTERS

Nonparametric Bayesian predictive distributions for future order statistics

Richard A. Johnson[a,*], James W. Evans[b], David W. Green[b]

[a] Department of Satistics, University of Wisconsin-Madison. Madison. WI 53706, USA
[b] USDA Forest Service, Forest Products Laboratory Madison. WI 53705, USA

Abstract

We derive the predictive distribution for a specified order statistic, determined from a future random sample, under a Dirichlet process prior. Two variants of the approach are treated and some limiting cases studied. A practical application to monitoring the strength of lumber is discussed including choices of prior expectation. and comparisons made to a Bayesian parametric approach. © 1999 Published by Elsevier Science B.V. All rights resewed

Keywords: Dirichlet process; Predictive distributions; Monitoring quality

1. Introduction

Many modem standards for the strength of materials or life-length of a product are specified in terms of a lower population percentile rather than a mean. In a nonparametric setting, an order statistic becomes the natural choice for a point estimate or one-sided confidence limit. The Bayesian nonparametric approach in this paper was motivated by talks with V. Susarla and his research including the estimation of survival curves in Susarla and Van Ryzin (1976) and the estimation of a branching process life-length distribution in Johnson et al. (1979).

The United States lumber industry sets standards in terms of a lower population percentile. This industry collected large samples of several strength properties on several different species and grades. These data can be used to construct informative prior distributions. Johnson et al. (1995) discuss this approach for normal. log-normal, and Weibull populations. They also calculate the posterior distributions of population percentiles. Concerning future data, predictive distributions can be employed to check conformity of future data with the informative prior distribution. Box (1980, 1981, 1983) develops this approach in the context of parametric families. In a nonparametric setting, we propose to monitor the strength of future specimens by determining whether or not the value of a future sample lower order statistic is reasonable in light of the informative prior.

* Corresponding author. E-mail: rich@stat.wisc.edu.

PII:S0l67-7152(98)00161-8

For application to the lumber problem, we derive the predictive distribution of future order statistics when (a) the preliminary data is used to develop an informative Dirichlet process prior and when (b) the preliminary data is combined with a possibly noninformative prior to obtain a posterior process. We also consider the limiting predictive distribution as the future sample size, or prior sample size, tends to infinity. This work complements the Bayesian parametric posterior and predictive distributions for order statistics derived in Johnson et al. (1995). Finally, we show how our results apply to monitoring the strength of lumber and compare the parametric and nonparametric Bayesian approaches.

2. Posterior distributions fur population percentiles and predictive distributions

The cumulative distribution function (cdf) of an order statistic is a function of the population cdf $F(\cdot)$. Employing a Dirichlet process prior, we derive the predictive distribution for a specified order statistic determined from a future random sample.

The nonparametric procedure requires that we put a prior directly on the cdf, $F(\cdot)$, of the strength property Ferguson (1973) describes the necessary mathematical structure for accomplishing this. In any application, is necessary to specify the prior expected probability that a single observation, Y does not exceed y. Denote this expected probability by $F_0(y)$. Continuing (Ferguson, 1973), we obtain a prior process which has

$$\text{mean} = F_0(y) \text{ for any } y,$$

$$\text{variance} = \frac{F_0(y)[1 - F_0(y)]}{c_0+1}$$

$$\text{covariance} = \frac{F_0(y_1 \wedge y_2)[- F_0(y_1 \vee y_2)]}{c_0+1} \tag{2.1}$$

Notice that the covariance structure is quite restrictive in form and that the variation depends on the single choice c_0 for all y.

The main advantage of selecting a Dirichlet process prior is that the posterior distribution of $F(\cdot)$ is again a Dirichlet process. When data $X_1 = x_1$, $X_2 = x_2, \ldots, X_m = x_m$ become available, the prior shape measure $c_0 F_0(\cdot)$ is up&ted to the measure

$$v_m(y) \equiv v_m(-\infty, y] = c_0 F_0(y) + (\text{number of } x_1, x_2, \ldots, x_m \leqslant y) \tag{2.2}$$

and the posterior expectation becomes $v_m(y)/(c_0 + m)$ or

$$\frac{c_0}{c_0+m}F_0 + \frac{m}{c_0+m}F_m \tag{2.3}$$

where $F_m(\cdot)$ is the empirical cdf of $x_1, x_2, \ldots, x_m$

Let h_q denote the $100q$th population percentile so $q = F(h_q)$ where, a posteriori, $F(\cdot)$ is distributed as a Dirichlet process with expectation measure (2.3). Ferguson (1973, Eq. (12), Section 5) gives an expression for the posterior distribution of a population percentile.

$$P[\eta_q \leqslant t] = P[F(t) > q]$$

$$= \int_q^1 \frac{\Gamma(c_0+m)}{\Gamma(v_m(t))\Gamma((c_0+m) - v_m(t))} u^{v_m(t)-1}(1-u)^{c_0+m-v_m(t)-1}\, du. \tag{2.4}$$

A 95% Bayesian lower tolerance bound for h_q is given by the largest value of t for which the right-hand side of Eq. (2.4) is less than or equal to 0.05. This value, $t = \underline{h}_q$ must be detemined numerically using an incomplete beta routine.

Let $Y_1,...,Y_n$ be a future random sample from $F(\cdot)$ and let $Y_{(r)}$ be the rth order statistic. Further, let P(F) be the Dirichlet process prior distribution for F(·) defined in terms of the measure $c_0\,F_0(\cdot)$. A future sample of size n will be selected. Given $F(\cdot)$, the cdf of the rth order statistic is given by

$$F_r(y) = \sum_{j=r}^{n} \binom{n}{j} F^j(y)[1 - F(y)]^{n-j}.$$

The predictive distribution is then

$$H_r = \int F_r \,\mathrm{d}P(F)$$
$$= \sum_{j=r}^{n} \binom{n}{j} \int F^j[1 - F]^{n-j}\,\mathrm{d}P(F).$$

2.1. Predictice distribution of $Y_{(r)}$ based on the prior process

For y fixed, $F(y)$ is distributed as a Beta[$c_0\,F_0(y)$. $c_0(1. - F_0(y))$] random variable. Consequently,

$$H_r(y) = \sum_{j=r}^{n} \binom{n}{j} \int_0^1 t^j(1-t)^{n-j}\frac{1}{B}\,t^{c_0F_0(y)-1}(1-t)^{c_0-c_0F_0(y)-1}\,\mathrm{d}t$$
$$= \sum_{j=r}^{n} \binom{n}{j} \int_0^1 \frac{1}{B}\,t^{c_0F_0(y)+j-1}(1-t)^{c_0-c_0F_0(y)+n-j-1}\,\mathrm{d}t$$
$$= \sum_{j=r}^{n} \binom{n}{j} \frac{B[c_0F_0(y)+j, c_0 - c_0F_0(y)+n-j]}{B[c_0F_0(y), c_0 - c_0F_0(y)]} \tag{2.5}$$

where $B[\cdot\,,\,\cdot]$ is the beta function.

If the prior expectation $F_0(\cdot)$ has a probability density function, $f_0(\cdot)$, the predictive distribution of an order statistic also has a probability density function. To obtain the predictive cdf, we first express each beta function in terms of gamma functions and these in terms of $\exp(\ln\Gamma(\cdot))$. For instance,

$$B[c_0F_0(y), c_0 - c_0F_0(y)] = \exp(\ln\Gamma[c_0F_0(y)] + \ln\Gamma[c_0(1-F_0(y))] - \ln\Gamma[c_0]).$$

Differentiating $H_r(y)$ with respect to y, we obtain an expression for the predictive pdf in terms of the psi function $\psi(z) = (\mathrm{d}/\mathrm{d}z)\ln\Gamma(z)$. In particular,

$$h_r(y) = \sum_{j=0}^{r-1} \binom{n}{j} \frac{B[c_0F_0(y)+j, c_0 - c_0F_0(y)+n-j]}{B[c_0F_0(y), c_0 - c_0F_0(y)]} c_0 f_0(y)$$
$$\times[\psi(c_0 - c_0F_0(y)+n-j) - \psi(c_0 - c_0F_0(y)) - \psi(c_0F_0(y)+j) + \psi(c_0F_0(y))]. \tag{2.6}$$

This expression. which is most convenient for small r is actually obtained by differentiating 1 - $H_r(y)$.

To see how the predictive distribution depends on c_0 and the future sample size n, we consider two limits. If $r = [np]$ and the future sample size grows unboundedly,

$$\lim_{n\to\infty} P[Y_{(r),n} \leqslant y \mid F] = 0(1) \quad \text{if} \quad F(y) < p \quad (F(y) > p)$$

so

$$\lim_{n\to\infty} P[Y_{(r),n} \leqslant y] = \lim_{n\to\infty} \int P[Y_{(r),n} \leqslant y \mid F]\,\mathrm{d}P(F)$$

$$= \int_p^1 \frac{1}{B[c_0F_0(y), c_0 - c_0F_0(y)]} t^{c_0F_0(y)-1}(1-t)^{c_0-c_0F_0(y)-1}\,\mathrm{d}t. \tag{2.7}$$

This distribution is not degenerate because of variability in the prior process.

Alternatively, we can consider the limit $c_0 \to \infty$. Fixing y, the limiting marginal prior distribution degenerate at $F_0(y)$ as long as $0 < F_0(y) < 1$. By the dominated convergence theorem, the limit distribution.

$$\lim_{c_0\to\infty} P[Y_{(r),n} \leqslant y] = \sum_{j=r}^{n} \binom{n}{j} F_0^j(y)[1-F_0(y)]^{n-j} \tag{2.8}$$

which is the usual distribution for an order statistic when the underlying cdf is $F_0(\cdot)$.

2.2. Predictive distributions based on posterior distributions

If preliminary data have already been used to update the prior so that the posterior distribution has expectation (2.3), we still proceed as above. Now, however, for a fixed y, the posterior distribution of $F(y) \mid x_1 \ldots .x$ is

$$\mathrm{Beta}[c_0F_0(y) + mF_m(y), c_0(1-F_0(y)) + m(1-F_m(y))].$$

Consequently, the predictive distribution of $Y_{(r).n}$ becomes

$$H_r(y) = \sum_{j=r}^{n} \binom{n}{j} \frac{B[c_0F_0(y) + mF_m(y) + j, c_0(1-F_0(y)) + m(1-F_m(y)) + n - j]}{B[c_0F_0(y) + mF_m(y), c_0(1-F_0(y)) + m(1-F_m(y))]}. \tag{2.9}$$

where $B[\cdot, \cdot]$ is the beta function.

The limits are similar to those above. In addition, if $m \to \infty$, and the true cdf $F_T(\cdot)$ is continuous at and $0 < F_T(y) < 1$, then

$$\lim_{m\to\infty} P[Y_{(r),n} \leqslant y] = \sum_{j=r}^{n} \binom{n}{j} F_T^j(y)[1-F_T(y)]^{n-j}. \tag{2.10}$$

3. An application of predictive distributions to monitor lumber

Following the approach developed by Box (1980, 1981, 1983) for parametric distributions, we propose employing the predictive distribution of an order statistic to check conformity of future data with the informative prior distribution. If the future population has `shifted downward. an order statistic based on a random sample should be in the extreme lower tail of the predictive distribution thus signaling a change.

We first determine an informative prior based on data collected a few years ago. For a fixed load. y, it is necessary to specify the prior expected probability that the strength of a single specimen, Y, does not exceed y. Denote this expected probability by $F_0(y)$. We then specify the uncertainty in the probability of failure $P[Y \leqslant y]$ as a beta distribution with

$$\mathbf{mean} = F_0(y)$$

and

$$\mathbf{variance} = \frac{F_0(y)[1-F_0(y)]}{c_0+1}.$$

That is. the probability of failure at load y_1 has probability density

$$g(u)=\frac{1}{B[c_0F_0(y_1),c_0(1-F_0(y_1))]}u^{c_0F_0(y_1)-1}(1-u)^{c_0(1-F_0(y_1))-1}. \tag{3.1}$$

For consistency of our specification at other y, we require that the expected probability $F_0(y)$ to be a cdf. Next. we must specify the probability distribution for two loads $y_1 < y_2$. The joint pdf of $P[Y \leqslant y_1] = u_1$ and $P[Y \leqslant y_2] = u_2$ is given by the Dirichlet distribution

$$\frac{u_1^{c_0F_0(y_1)-1}(u_2-u_1)^{c_0(F_0(y_2)-F_0(y_1))-1}(1-u_2)^{c_0(1-F_0(y_2))-1}}{D[c_0F_0(y_1),c_0(F_0(y_2)-F_0(y_1)),c_0(1-F_0(y_2))]}.$$

Continuing to define higher dimensional densities for the probabilities at loads $y_1 < y_2 < \cdots < y_n$ we obtain a Dirichlet process prior distribution on the cdf $F(\cdot)$ with

mean $F_0(y)$ for any y.

$$\text{variance}=\frac{F_0(y)[1-F_0(y)]}{c_0+1},$$

$$\text{covariance}=\frac{F_0(y_i \wedge y_j)[1-F_0(y_i \vee y_j)]}{c_0+1}. \tag{3.2}$$

We have relatively large data sets $x_1,\ldots,x_{n0}$ from which to develop a prior process. We took $c_0 = n_0 - 1$ although other choices are possible. (If no prior observations are available then $F_0(y)[1 - F_0(y)]in_0$ is the sampling variance of the fraction that fail at load y.)

Based on fits to $x_1,\ldots,x_{n0}$, for one grade and species, our two choices for $F_0(\cdot)$ were the continuous distributions

Normal with mean = 9.870 and s.d. = 2.490.
Weibull with shape = 4.726 and scale 10.803.

The predictive cdf's for the sample 5th 10th and 50th percentiles were evaluated by expressing the beta functions in terms of gamma functions. The cdf's were evaluated on a grid of points in steps of 0.001. Some of the percentiles were checked with a double precision calculation and they were found to be accurate within the step size 0.001 except for some of the 1st percentiles which were occasionally off by 0.004 units.

A future sample of size n will be taken and the rth order statistic obtained. If this is too small relative to its predictive distribution, a decrease in strength is suggested. We selected the lower 5th percentile of the predictive distribution as the critical value for signaling a decrease in strength. One approach to evaluating this monitoring procedure is to determine the probability that it signals a change when the prior. process expectation F_0 is changed to an alternative distribution F_{0A}. Under this alternative prior Dirichlet process, a cdf is selected and then a sample of size n is taken. The sample rth order statistic $Y_{(r)}$ has distribution H_{rA} where, as in Eq. (2.5) with F_0 replaced by F_{0A},

$$H_{rA}(y)=\sum_{j=r}^{n}\binom{n}{j}\frac{B[c_0F_{0A}(y)+j,c_0-c_0F_{0A}(y)+n-j]}{B[c_0F_{0A}(y),c_0-c_0F_{0A}(y)]}.$$

A change to a weaker strength will be signaled if $Y_{(r)}$ is smaller than the 5th percentile $\xi_{0.05}$ where $H_r(\xi 0.05) = 0.05$ with $H_r(\cdot)$ given by Eq. (2.5).

The probability of signaling a decrease in strength is then

$$H_{rA}(\xi_{0.05})=\sum_{j=r}^{n}\binom{n}{j}\frac{B[c_0F_{0A}(\xi_{0.05})+j,c_0-c_0F_{0A}(\xi_{0.05})+n-j]}{B[c_0F_{0A}(\xi_{0.05}),c_0-c_0F_{0A}(\xi_{0.05})]}. \tag{3.3}$$

Table 1
Detection probabilities: nonparametric Bayes predictive distributions for the sample 5th percentile (shifted predictive distribution)

sample size	Shift in predictive distribution								
	0.000	0.250	0.500	0.750	1.000	1.250	1.500	1.750	2.000
Prior size = 120									
51	0.050	0.077	0.116	0.172	0.247	0.342	0.454	0.576	0.695
201	0.050	0.092	0.164	0.275	0.425	0.599	0.764	0.888	0.959
351	0.050	0.096	0.176	0.302	0.471	0.659	0.823	0.930	0.980
501	0.050	0.099	0.186	0.322	0.505	0.700	0.859	0.952	0.989
Prior size = 360									
51	0.050	0.084	0.135	0.208	0.306	0.425	0.555	0.685	0.799
201	0.050	0.120	0.252	0.447	0.668	0.848	0.951	0.989	0.999
351	0.050	0.135	0.305	0.548	0.786	0.934	0.988	0.999	1.000
501	0.050	0.148	0.346	0.619	0.854	0.968	0.996	1.000	1.000
Prior size = 1000									
51	0.050	0.087	0.144	0.226	0.333	0.460	0.597	0.726	0.834
201	0.050	0.141	0.315	0.557	0.786	0.929	0.985	0.998	1.000
351	0.050	0.173	0.421	0.720	0.919	0.988	0.999	1.000	1.000
501	0.050	0.200	0.507	0.822	0.969	0.998	1.000	1.000	1.000

Note: The Dirichlet prior has a normal distribution, with mean = 9.87 and s.d. = 2.49, as its mean.

By selecting the future sample directly from F_{0A}, we obtain a second measure of the probability of signaling a change which is expressed in terms of the cdf $G_{r,0,A}$ of $Y_{(r)}$.

$$G_{r,0,A}(\xi_{0.05}) = \sum_{j=r}^{n} \binom{n}{j} F_{0A}^{j}(\xi_{0.05})[1 - F_{0A}(\xi_{0.05})]^{n-j}. \tag{3.4}$$

In order to study the power of the predictive distributions for detecting changes in the strength population, we considered shifts in the population of 0.000(0.250)2.000 units. Taking a cdf from the original process and shifting it by the specified amount is equivalent to shifting the prior expectation F_0 by the same amount and then selecting a cdf from this modified process. Tables 1 and 2 pertain to alternatives that contain two sources of variation. First the choice of a cumulative distribution function according to the informative prior which is then shifted downward by the specified amount of shift Second the sampling variation of the order statistic, representing the sample 5th percentile, obtained from a random sample of size n from this shifted distribution function. Equivalently, the predictive distribution of the order statistic can be shifted.

Tables 1 and 2 give values of the detection probability which is the predictive probability, under the shifted predictive distribution, that lies below the 5th percentile of the un-shifted predictive distribution of the order statistic.

In review, suppose the process generating the new observations tends to produce values D units below those suggested by current data. Tables 1 and 2 provide the amount of predictive probability, calculated from the new data, that lies below the 5th percentile signal value determined from the predictive distribution based on the current study data. Inspection of Tables 1 and 2 indicates that, generally, shifts must be greater than 0.750 in order to have predictive probabilities around 0.5 or larger. The exceptions are for the smallest sample size, n = 51, where the shifts must be about 1.000.

Table 2
Detection probabilities: nonparametric Bayes predictive distributions for the sample 5th percentile (shifted predictive distribution)

Future sample size	Shift in predictive distribution								
	0.000	0.250	0.500	0.150	1.000	1.250	1.500	1.750	2.000
Prior size = 120									
51	0.050	0.078	0.119	0.176	0.252	0.346	0.455	0.572	0.686
201	0.050	0.093	0.164	0.272	0.417	0.583	0.744	0.869	0.946
351	0.050	0.096	0.176	0.298	0.460	0.641	0.803	0.915	0.972
501	0.050	0.099	0.184	0.317	0.491	0.679	0.838	0.938	0.983
Prior size = 360									
51	0.050	0.085	0.138	0.212	0.309	0.424	0.550	0.675	0.785
201	0.050	0.119	0.246	0.434	0.646	0.827	0.937	0.984	0.997
351	0.050	0.134	0.297	0.531	0.764	0.918	0.982	0.996	1.000
501	0.050	0.146	0.336	0.598	0.832	0.956	0.994	1400	1.000
Prior size = 1000									
51	0.050	0.088	0.146	0.229	0.334	0.458	0.589	0.714	0.819
201	0.050	0.139	0.306	0.522	0.763	0.913	0.978	0.997	1.000
351	0.050	0.171	0.408	0.698	0.903	0.982	0.998	1.000	1.000
501	0.050	0.196	0.488	0.799	0.959	0.996	1.000	1.000	1.000

Note: The Dirichlet prior has a Weibull distribution, with shape = 4.726 and scale = 10.803, as its mean.

4. Comparison of the Bayesian nonparametric and parametric approaches

Johnson et al. (1995) take a parametric approach. They assume either a two parameter Weibull or a normal distribution for a strength property. In the normal case, values for the population mean and variance are selected from an informative prior distribution and a random sample of size n generated from the corresponding normal distribution. The sample percentile obtained is thus an order statistic from a random sample of normal variables, given the values for the population mean and variance. The detection probabilities for normal and Weibull populations are presented in their Tables 27.1 and 27.2, and Tables 27.5 and 27.6, respectively.

In the nonparametric approach. a population cdf $F(\cdot)$ is selected from the informative prior. This cdf will likely be close to the normal distribution that is the expected value of the Dirichlet process but it will not be normal. A random sample of size n is then selected From $F(\cdot)$ and the order statistic obtained.

The major difference in the two procedures is that the nonparametric Bayesian approach does not require that the future random sample be generated from a normal distribution. It allows more latitude in the choice of the cdf $F(\cdot)$ for a strength property. Heuristically, this should imply more variability in the population percentile and hence in the sample percentile.

More importantly, suppose the underlying distribution $F(t)$ changes for $t < t_0$ but remains unaltered for $t \geqslant t_0$ for some t_0. This kind of change cannot be incorporated in common parametric models but is easily accommodated in a Dirichlet process distribution of F by an appropriate change in F_0 for $t < t_0$.

Comparing our Table 1 with Tables 27.1 and 27.2 from Johnson et al. (1995) and our Table 2 with Tables 27.5 and 27.6, we note that the estimators of the 5th percentile are ordered, by increasing detection probabilities.

Bayesian nonparametric (lowest).
Order statistic from parametric distribution.
Parametric estimator (highest).

This ordering holds for both the Weibull and normal cases.

Once it is decided **that** it is necessary to detect a shift in the 5th percentile of a given magnitude. the calculations leading to Table 1 can provide a means for making an intelligent selection of future sample size.

References

Box. G.. 1980. Sampling and Bayes' inference in the advancement of learning. In: De Groot. M.H.. Benardo. J.M.. Lindley. D.V.. Smith A.F.M. (Eds.). Bayesian Statistics. University Press. Valencia. pp. 366–381.

Box. G.. 1981. Sampling and Bayes' inference in scientific modeling and robustness. J. Roy. Statist. Soc.. A 143. 383–430 (with discussion).

Box. G.. 1983. An apology for ecumenism in statistics. In: Box. G.E.P.. Leonard. T.. Wu. C.-F. (Eds.). Scientific Inference. Data Analysis. and Robustness. Academic Press. New York. pp. 51–84.

Ferguson. T.. 1973. Bayesian analysis of some nonparametric problems. Ann. Statist. 1. 209–230.

Johnson. R.. Evans. J.. Green. D.. 1995. Predictive distributions of order statistics. In: Balakrishnan. N. (Ed.). Recent Advances in Life-Testing and Reliability. CRC Press. Boca Raton. pp. 505–521.

Johnson. R., Susarla. V.. Van Ryzin. J.. 1979. Bayesian non-parametric estimation for age-dependent branching processes. Stochastic Process. Appl. 9. 301–318.

Susarla. V.. Van Ryzin. J.. 1976. Nonparametric Bayesian estimation of survival curves from incomplete observations. J. Amer. Statist. Assoc. 61. 897–902.