STATISTICAL RESEARCH REPORT No 12
Institute of Mathematics 1971
University of Oslo

ON PREDICTION REGIONS

by

Emil Spjøtvoll

ABSTRACT

A theorem on the construction of smallest volume invariant prediction regions for the observed values of random variables is given. Various applications of the theorem is given, including predictions from a regression function and predictions of the values of the observations from an exponential distribution. It is shown by a counterexample that smallest volume invariant prediction regions do not necessarily minimize volume uniformly in the class of all prediction regions.

1. Introduction. Let the random variable Y over the measurable space $(\mathcal{Y}, \mathcal{B})$ have a density $h(y)$ with respect to a measure ν. We shall consider the problem to predict the observed value y of Y before y is observed, and shall construct a region, a prediction region, which has a given probability $1-\alpha$ of covering the observed value. Such a region is indicated by a measurable function $\varphi(y)$, where $\varphi(y)$ is the probability that y is in the prediction region. Let ν^* be a measure defined on $\mathcal{Y}$ which is used to define volume in the space $\mathcal{Y}$ so that the volume of a set $B \in \mathcal{B}$ is

$$\nu^*(B) = \int_B d\nu^*(y) .$$

We shall assume that ν^* is absolutely continuous with respect to ν, and that $d\nu^*(y) = a(y)\, d\nu(y)$. Hence

$$\nu^*(B) = \int_B a(y)\, d\nu(y) .$$

The volume of the prediction region given by the prediction function $\varphi(y)$ is then

$$\int \varphi(y)\, a(y)\, d\nu(y) .$$

Definition. *The prediction function* $\varphi(y)$ *is a smallest volume prediction function for* Y *at prediction level* $1-\alpha$ *if*

(1) $$\int \varphi(y)\, h(y)\, d\nu(y) \geq 1-\alpha ,$$

and for any other prediction function ψ *satisfying*

$$\int \psi(y)\, h(y)\, d\nu(y) \geq 1-\alpha$$

we have

$$\int \psi(y)\, a(y)\, d\nu(y) \geq \int \varphi(y)\, a(y)\, d\nu(y) .$$

A prediction region corresponding to a $\varphi(y)$ satisfying (1) is by Guttman [8, p. 5] called a $(1-\alpha)$-expectation tolerance region. Fraser and Guttman [4], Guttman [6], [7] and [8] considered another optimality property of prediction regions than smallest volume by introducing a weighing function indicating the desirability of the various regions.

In Section 2 we shall need the following

Lemma. *The smallest volume prediction level* $1-\alpha$ *is of the form*

$$\varphi(y) = \begin{cases} 1 & \text{when} \quad a(y) < kh(y) \\ 0 & \text{when} \quad a(y) > kh(y) \end{cases}$$

where k *and the value of* $\varphi(y)$ *when* $a(y) = kh(y)$ *are determined so that* $\int \varphi(y)\, h(y)\, d\nu(y) = 1-\alpha$.

Proof. Follows easily by an application of the Neyman-Pearson lemma.

Suppose now that the density of Y depends upon an unknown parameter θ, the density being $h(y,\theta)$, $\theta \in \Omega$. Before predicting the value of Y we observe the value of a random variable X independent of Y which has density $f(x,\theta)$ w.r.t. a σ-finite measure μ on a measurable space $(\mathcal{X}, \mathcal{A})$. We can now indicate a prediction region for Y by a measurable function $\varphi(x,y)$ which is the probability that y is included in the prediction region when X has the value x. When $X = x$, the volume of the predicted region is

$$\int \varphi(x,y)\, d\nu^*(y)$$

The unconditional volume is

$$\int [\int \varphi(x,y)\, d\nu^*(y)] f(x,\theta)\, d\mu(x) .$$

The prediction level of the prediction function $\varphi(x,y)$ is

$$E_\theta\ \varphi(X,Y)\ .$$

The prediction region has prediction level $1-\alpha$ if

$$E_\theta\ \varphi(X,Y) \geq 1-\alpha \qquad \text{for all } \theta.$$

2. The theorem. In the following we shall assume that there exists a group G of transformation g of (X,Y) such that the distribution of $g(X,Y)$ belongs to the family of distributions of (X,Y). Let $\bar{g}\theta \in \Omega$ be the parameter corresponding to the distribution of $g(X,Y)$ when θ is the parameter of the distribution of (X,Y). Assume also that $\bar{g}\Omega = \Omega$. Let $\bar{G}$ be the group of all $\bar{g}$. If the assumptions above hold we shall say that the prediction problem is invariant under the group G.

A prediction function $\varphi(x,y)$ is invariant if

$$\varphi(g(x,y)) = \varphi(x,y) \qquad \text{for all } x,\ y \text{ and } g\ .$$

We shall consider invariant prediction functions. Let $T(x,y)$ be a maximal invariant function under G. An invariant prediction function depends only upon $T(x,y)$. For the concepts used here, see, e.g., Lehmann [9].

Suppose that $\bar{G}$ is transitive over Ω. Then the distribution of $T(X,Y)$ does not depend upon θ. Let $T(X,Y)$ have density $p(t)$ w.r.t. a measure λ over the space $(\mathcal{T},\mathcal{B})$. Let λ^* be a measure which indicates volume in $\mathcal{T}$, and let $d\lambda^*(t)/d\lambda(t) = c(t)$. We make the following

Assumption. Let $T_x(y) = T(x,y)$. Then $\nu^* T_x^{-1} = \lambda^*$ for all x.

<u>Theorem.</u> <u>Let</u> $\psi(t)$ <u>be defined by</u>

$$\psi(t) = \begin{cases} 1 & \underline{\text{when}} \quad c(t) < kp(t) \\ \gamma & \underline{\text{when}} \quad c(t) = kp(t) \\ 0 & \underline{\text{when}} \quad c(t) > kp(t) \end{cases}$$

<u>where</u> k <u>and</u> γ <u>are determined so that</u> $E\ \psi(T) = 1-\alpha$. <u>Then</u> $\varphi(x,y) = \psi(T(x,y))$ <u>is a smallest volume invariant prediction function at prediction level</u> $1-\alpha$.

<u>Proof.</u> We have

$$E\ \varphi(X,Y) = E\ \psi(T) = 1-\alpha\ .$$

Hence φ has prediction level $1-\alpha$. Let $\varphi_1(x,y)$ be any other invariant prediction function at level $1-\alpha$. Since φ_1 is invariant it depends on x and y only through T . Hence $\varphi_1(x,y) = \psi_1(T(x,y))$. Since by the Lemma, ψ minimizes

$$\int \varphi(t)\ d\lambda^*(t)$$

among all prediction functions φ , we must have

$$\int \psi_1(t)\ d\lambda^*(t) \geq \int \psi(t)\ d\lambda^*(t)\ .$$

Then by the Assumption

$$\int \varphi_1(x,y)\ d\nu^*(y) \geq \int \varphi(x,y)\ d\nu^*(y)$$

for all x . Hence

$$\int\left[\int \varphi_1(x,y)\ d\nu^*(y)\right] f(x,0)d\mu(x) \geq \int\left[\int \varphi(x,y)d\nu^*(y)\right] f(x,0)d\mu(x)\ ,$$

which shows that the expected volume using φ_1 is not smaller than the expected volume using φ .

3. Examples. Example 1. Let $X_1,\dots,X_n$ be independent $N(\mu,\sigma^2)$ where μ and σ^2 are unknown. Consider the problem to predict the values of q independent random variable $Y_1,\dots,Y_q$ also from a $N(\mu,\sigma^2)$ distribution. Since

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

are sufficient we shall base the prediction on $\bar{X}$ and S^2. The problem is invariant under the transformations

$$\bar{X}' = a\bar{X} + b$$

$$S' = aS$$

$$Y_i = aY_i + b \qquad i = 1,\dots,q$$

where $a > 0$ and b are real numbers. Maximal invariant are

$$T_i = \frac{Y_i-\bar{X}}{S} \qquad i = 1,\dots,q\ .$$

The variable $T' = (T_1,\dots,T_q)$ has a multivariate t-distribution with $n - 1$ degrees of freedom and covariance matrix $I + n^{-1}E$ (see [2]), where I is the identity matrix and E is the matrix with all elements equal to one. The density of T is proportional to

$$\{1 + t'(I+n^{-1}E)^{-1}t/(n+1)\}^{-\frac{1}{2}(n+q-1)}$$

Let both λ and λ^* be the Lebesgue measure in q-dimensional Euclidean space. Then the function $\psi(t)$ in the Theorem is of the form

$$\psi(t) = \begin{cases} 1 & \text{when } t'(I+n^{-1}E)^{-1}t \le C \\ 0 & \text{when } t'(I+n^{-1}E)^{-1}t > C\ , \end{cases}$$

where C is a constant. The statistic

$$\frac{1}{q}\,T'(I+n^{-1}E)^{-1}T$$

has an F-distribution with q and $n-1$ degrees of freedom. To get a $1-\alpha$ prediction region we therefore have to choose

$C = qF_{\alpha,q,n-1}$, where $F_{\alpha,q,n-1}$ is the upper α-point of the F-distribution with q and $n-1$ degrees of freedom. Using the fact that $(I + n^{-1}E)^{-1} = I - (n+q)^{-1}E$ it is found that the corresponding prediction region for $Y_1,\ldots,Y_q$ is

$$(2) \quad \{(y_1,\ldots,y_q) : \sum_{i=1}^{q} (y_i-\bar{x})^2 - q^2(n+q)^{-1}(\bar{y}-\bar{x})^2 \leq qs^2F_{\alpha,q,n-1}\}$$

The assumption is satisfied with ν^* the Lebesgue measure. We then have by the definition of t that $t \in C_2 \subset \mathscr{E}$ if and only if $y \in Cs + \bar{x}$, and $\nu^* T_x^{-1}(C) = \nu^*(Cs + \bar{x}) = \nu^*(C)$ which is equal to $\lambda^*(C)$. It follows that (2) gives the smalles volume invariant prediction region for $Y_1,\ldots,Y_q$.

A derivation of optimal region in the Fraser and Guttman sense for this problem can (for the case $q = 1$) be found in [4], [6] and [8]. A generalization to the multivariate case will not bring any theoretical difficulties. The results for $q = 1$ will be the same as those in [1], [4] and [8].

Example 2. Consider a regression problem where the random variables $X_1,\ldots,X_n$ are independent, normal with

$$\begin{aligned} EX_i &= \sum_{j=1}^{p} z_{ji}\beta_j \\ & \qquad\qquad\qquad i = 1,\ldots,n , \\ \operatorname{Var} X_i &= \sigma^2 \end{aligned}$$

where the $\{z_{ij}\}$ are known constants and $\beta_1,\ldots,\beta_p,\sigma^2$ are unknown parameters. Let Z be the matrix with elements $\{z_{ij}\}$, X the vector with elements X_i and β the vector with elements β_i . The least square estimate of β is

$$\hat{\beta} = (Z\,Z')^{-1}Z\,X .$$

The estimate $\hat{\beta}$ is $N(\beta, A\sigma^2)$, where $A = (Z Z')^{-1}$.
The statistics $\hat{\beta}$ and

$$S^2 = \frac{1}{n-p} (X-Z'\hat{\beta})'(X-Z'\hat{\beta})$$

are sufficient, and we shall base our prediction procedure on these.

We shall consider the problem to predict the values of q future dependent variables $Y_1,\ldots,Y_q$, where Y_i is the value of the dependent variable when the independent variables have the values $(v_{1i},\ldots,v_{pi}) = v_i'$, $i=1,\ldots,q$. Then Y_i is $N(v_i' \beta,\sigma^2)$, $i=1,\ldots,q$. The prediction problem is invariant under the transformations

$$\hat{\beta}_i' = a(\hat{\beta}_i + b_i) \qquad i = 1,\ldots,p$$

$$S' = aS$$

$$Y_i' = a(Y_i + \sum_{j=1}^{p} v_{ji}b_j) \qquad i = 1,\ldots,q .$$

Maximal invariant are

$$T_i = \frac{Y_i - \Sigma_{j=1}^{q} v_{ji}\hat{\beta}_j}{S} \qquad i = 1,\ldots,q .$$

The random vector $T' = (T_1,\ldots,T_q)$ has a multivariate t-distribution with $n-p$ degrees of freedom and covariance matrix $I + V'AV$ where $V = [v_1,\ldots,v_q]$. Using the same ν^* and λ^* as in Example 1 we find that the smallest volume level $1-\alpha$ invariant prediction region is

$$\{y : (y-V'\hat{\beta})'(V'AV)^{-1}(y-V'\hat{\beta}) \leq qs^2F_{\alpha,q,n-p}\}. \tag{3}$$

Lieberman [10] compared the prediction regions given by (3) and prediction regions obtained by a simple application of separate t-statistics in the case $q=2$, and found that the latter gives smaller prediction intervals. But if we consider the joint prediction region it follows from the Theorem, since both regions are invariant, that (3) gives a region with smaller volume. A discussion of the two methods of constructing prediction region can also be found in Miller [11, pp. 114-116].

Example 3. Let $X_1,\dots,X_n$ be n independent random variables from an exponential distribution

$$(4) \qquad \theta e^{-\theta x} \qquad x > 0 .$$

Let the problem be to predict the values of q independent random variables $Y_1,\dots,Y_q$, also with the density (4). We shall base the prediction on the sufficient statistic $\overline{X} = n^{-1} \Sigma_{i=1}^{n} X_i$. The problem is invariant under the transformations

$$\overline{X}' = a\,\overline{X}$$

$$Y_i' = a\,Y_i \qquad i=1,\dots,q ,$$

where $a > 0$. Maximal invariant are

$$T_i = \frac{Y_i}{n\overline{X}} \qquad i=1,\dots,q .$$

Using the fact that $2n\theta\overline{X}$ and $2\theta Y_i$ are independent and have chr-square distributions with $2n$ and 2 degrees of freedom, respectively, it is easily found that the joint density of $T_1,\dots,T_q$ is proportional to

$$\Big(\sum_{i=1}^{q} t_i + 1\Big)^{-(n+q)} .$$

Again, letting λ and λ^* be the Lebesgue measure we find that

$$\psi(t) = \begin{cases} 1 & \text{when } \Sigma_{i=1}^{q} t_i \le C \\ 0 & \text{when } \Sigma_{i=1}^{q} t_i > C , \end{cases}$$

where C is a constant. The variable

$$\frac{n}{q} \sum_{i=1}^{q} T_i = \frac{1}{q} \frac{\Sigma_{i=1}^{q} Y_i}{\Sigma_{i=1}^{n} X_i}$$

has an F-distribution with $2q$ and $2n$ degrees of freedom. It follows that $C = q\, F_{\alpha,2q,2n}/n$.

The prediction region for $Y_1,\dots,Y_q$ becomes

$$(5) \qquad \{(y_1,\dots,y_q) : \sum_{i=1}^{q} y_i \le q\bar{x}F_{\alpha,2q,2n}\}$$

The Assumption is easily seen to be satisfied with ν^* equal to the Lebesgue measure. Hence (5) is the smallest volume invariant prediction region.

For the problem of finding optimal prediction regions in the Fraser an Guttman sense for this problem see Guttman [7] and [8]. Goodman and Madansky [5] have studied the problem to find tolerance regions for an exponential distribution.

<u>Example 4.</u> As an extension of Example let $X_1,\dots,X_n$ have the density

$$(6) \qquad \theta e^{-\theta(x-\mu)} \qquad\qquad x > \mu ,$$

where both μ and θ are unknown. This example is also studied by Guttman [7] and [8]. The statistics $V = \min_i X_i$ and $W = \sum_{i=1}^{n} (X_i - V)$ are sufficient, and V and W are independent. $2n\theta V$ has a chi-square distribution with 2 degrees of freedom and $2\theta W$ has a chi-square distribution with $2(n-1)$ degrees of freedom, see [3].

Let $Y_1,\dots,Y_q$ from the distribution (6) be q variables to be predicted. The problem is invariant under the transformation.

$$V' = aV + b$$

$$W' = aW$$

$$Y_i' = aY_i + b \qquad\qquad i = 1,\dots,q$$

where $a > 0$. Maximal invariant is

$$T_i = \frac{Y_i - V}{W} \qquad i = 1,\ldots,q .$$

We shall study only the case $q = 1$. The density of $T = T_1$ is found to be

$$(7) \qquad p(t) = \begin{cases} \frac{n(n-1)}{n+1}\,(1+t)^{-n} & \text{when} \quad t > 0 \\ \frac{n(n-1)}{n+1}\,(1-nt)^{-n} & \text{when} \quad t < 0 . \end{cases}$$

From (7) and the Lemma we find that the $1-\alpha$ prediction region for T has the form

$$[-\tfrac{1}{n}(1 - \frac{1}{\alpha^{1/(n-1)}}) , \quad (1 - \frac{1}{\alpha^{1/(n-1)}})] .$$

The region for Y_1 then becomes

$$[v - \tfrac{1}{n}(1 - \frac{1}{\alpha^{1/(n-1)}})w , \quad v + \tfrac{1}{n}(1 - \frac{1}{\alpha^{1/(n-1)}})w] .$$

4.A counterexample. The following example shows that the smallest volume invariant prediction region does not necessarily have smallest volume in the class of all prediction region. Consider again Example 1 and let $q = 1$ and σ^2 be known and equal to 1. It is easily found that the smallest volume invariant prediction region for $Y = Y_1$ is

$$[\bar{x} - (1+n^{-1})^{\frac{1}{2}} z_{\alpha/2} , \quad \bar{x} + (1+n^{-1})^{\frac{1}{2}} z_{\alpha/2}]$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ point of standard normal distribution. The volume of this region is

$$(8) \qquad 2(1+n^{-1})^{\frac{1}{2}} z_{\alpha/2} .$$

Consider the following alternative prediction region for Y . If $\bar{x} < -\frac{1}{n+1}(1+\frac{1}{n})^{\frac{1}{2}} z_\alpha$, then use the region $[\bar{x} - (1+\frac{1}{n})^{\frac{1}{2}} z_\alpha, \; n\bar{x}]$.

If $-\frac{1}{n+1}(1+\frac{1}{n})^{\frac{1}{2}}z_\alpha \leq \bar{x} \leq \frac{1}{n+1}(1+\frac{1}{n})^{\frac{1}{2}}z_\alpha$, then use the region $[\bar{x}-(1+\frac{1}{n})^{\frac{1}{2}}z_\alpha , \bar{x}+(1+\frac{1}{n})^{\frac{1}{2}}z_\alpha]$. If $\bar{x} > \frac{1}{n+1}(1+\frac{1}{n})^{\frac{1}{2}}z_\alpha$, then use the region $[-n\bar{x} , \bar{x}+(1+\frac{1}{n})^{\frac{1}{2}}z_\alpha]$.

The region has prediction level $1-\alpha$ since we can verify that the probability that Y is in the prediction region is equal to

$$P[Y-\bar{X} \geq -(1+\tfrac{1}{n})^{\frac{1}{2}}z_\alpha \text{ and } Y+n\bar{X} < 0]$$

$$+ P[Y-\bar{X} \leq (1+\tfrac{1}{n})^{\frac{1}{2}}z_\alpha \text{ and } Y+n\bar{X} \geq 0]$$

$$= (1-\alpha)P[Y+n\bar{X} < 0] + (1-\alpha)P[Y+n\bar{X} \geq 0] = 1-\alpha .$$

When $-(n+1)^{-1}(1+n^{-1})^{\frac{1}{2}}z_\alpha \leq \bar{x} \leq (n+1)^{-1}(1+n^{-1})^{\frac{1}{2}}z_\alpha$, the volume of the prediction region is $2(1+n^{-1})^{\frac{1}{2}}z_\alpha$, which is smaller than (8) since $z_\alpha < z_{\alpha/2}$.

REFERENCES

[1] CHEW, V. (1966). Confidence, prediction, and tolerance regions for the multivariate normal distribution. J. Amer. Statist. Assoc. **61** 605-617.

[2] DUNNETT, C.W. and SOBEL, M. (1954). A bivariate generalization of Student's t-distribution, with tables for certain special cases. Biometrika **41** 153-169.

[3] EPSTEIN, B. and SOBEL, M. (1953). Life testing. J. Amer. Statist. Assoc. **48** 486-502.

[4] FRASER, D.A.S. and GUTTMAN, I. (1956). Tolerance regions. Ann. Math. Statist. **27** 169-179.

[5] GOODMAN, L.A. and MADANSKY, A. (1962). Parameter-free and nonparametric tolerance limits: the exponential case. Technometrics **4** 75-96.

[6] GUTTMAN, I. (1957). On the power of optimum tolerance regions when sampling from normal distributions. Ann. Math. Statist. **28** 773-778.

[7] GUTTMAN, I. (1959). Optimum tolerance regions and power when sampling from some non-normal universes. Ann. Math. Statist. **30** 926-938.

[8] GUTTMAN, I. (1970). Statistical Tolerance Regions: Classical and Baysian. No. 26 of Griffins Statistical Monographs & Courses. Griffin, London.

[9] LEHMANN, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

[10] LIEBERMAN, G.J. (1961). Prediction regions for several predictions from a single regression line. Technometrics **3** 21-27.

[11] MILLER, R.G. Jr. (1966). Simultaneous Statistical Inference. McGraw-Hill, New York.