İstatistik, Journal of the Turkish Statistical Association
2000, Vol. 3, No 3, pp. 58-82

PREDICTION INTERVALS FOR A DISCRETE EXPONENTIAL FAMILY OF DISTRIBUTIONS AND ITS APPLICATIONS

Masafumi Akahira and Eisuke Hida
Institute of Mathematics
University of Tsukuba
Ibaraki 305-8571, Japan

Abstract

Let **X** be an observable random vector and Y a random variable to be observed in future. Assume that the joint distribution of **X** and Y depends on an unknown parameter. In this paper we consider a way of the construction of a prediction interval for Y based on **X** for a discrete exponential family of distributions. In particular we asymptotically construct the prediction interval in the binomial and Poisson cases, and give practical applications to the prediction of the number of wins of the Japanese professional baseball teams and that of home runs of the players in the major league of the United States.

Key Words: (Similar) prediction region; prediction intervals; confidence coefficient; sufficient statistics; Cornish–Fisher expansion; binomial case; Poisson case.

1. Introduction

In a statistical inference, we may consider a predictive procedure for an unobserved random variable based on an observable random vector (see, e.g. Guttman(1970), Lauritzen(1974), Takeuchi(1975), Hinkley(1979), Butler(1986), Akahira(1990), Bjϕrnstad(1990), Geisser(1993), Takada(1996), Barndorff-Nielsen and Cox(1996)).

Suppose that $\mathbf{X} = (X_1, \cdots, X_m)$ is an observable random vector, Y is a random variable to be observed in future, and the joint distribution of $(\mathbf{X}, Y)$ depends on an unknown parameter θ in Θ, where Θ is a parameter space. Let $\mathcal{Y}$ be a space representing the possible outcomes of Y. If for any α $(0 < \alpha < 1)$ there exists a subset $S_{\mathbf{X}}$ (of $\mathcal{Y}$) based on $\mathbf{X}$ such that

$$P_\theta\{Y \in S_{\mathbf{X}}\} \geq 1 - \alpha, \qquad \text{for all } \theta \in \Theta, \tag{1}$$

then $S_{\mathbf{X}}$ is called a prediction region of Y at confidence coefficient $1-\alpha$. If $\mathcal{Y}$ is a subset of $\mathbf{R}^1$ and $S_{\mathbf{X}}$ is an interval $[a(\mathbf{X}), b(\mathbf{X})]$, then $S_{\mathbf{X}}$ is called a prediction interval of Y at confidence coefficient $1-\alpha$ (see Figure 1). If $\mathbf{X}$ takes a realized value $\boldsymbol{x} = (x_1, \cdots, x_m)$, then the interval $[a(\boldsymbol{x}), b(\boldsymbol{x})]$ is called a prediction interval of Y at confidence coefficient $100(1-\alpha)\%$. If, in particular, the equality in (1) holds, then the prediction region $S_{\mathbf{X}}$ is said to be similar.

In this paper we consider the case when the joint distribution of $(\mathbf{X}, Y)$ belongs to a discrete exponential family of distributions with an unknown one-dimensional parameter θ. Since there exists a complete and sufficient statistic T, using a conditional distribution of Y given T we obtain the conditional mean, variance and third cumulant, and give a way to construct a prediction interval of Y based on $\mathbf{X}$, by the Cornish–Fisher expansion. Indeed, for the binomial and Poisson cases, we asymptotically obtain the prediction intervals and curves for Y, and give practical applications to the prediction of the number of wins of the Japanese professional baseball teams and that of home runs of the players in the major league of the United States.

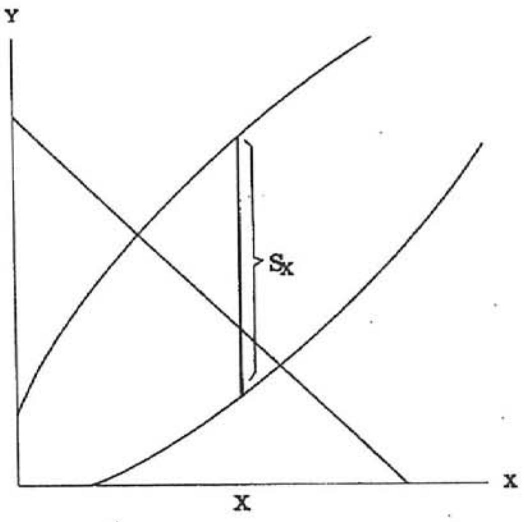

Figure 1: Prediction interval S_X of Y based on X

2. Prediction intervals for a discrete exponential family of distributions

Suppose that $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ are independent and identically distributed random variables according to a one-parameter exponential type distribution with a probability mass function (or p.m.f. for short)

$$f(x;\theta) = c(\theta)h(x)\exp\{\eta(\theta)t(x)\}$$

for $x = 0, 1, 2, \cdots, \theta \in \Theta = \mathbf{R}^1$, where $c(\theta)$ and $h(x)$ are nonnegative real-valued functions of θ and x, respectively, and $\eta(\theta)$ and $t(x)$ are real-valued functions of θ and x, respectively. Then the joint p.m.f. of $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ is given

$$f_{X_1,\cdots,X_m,Y_1,\cdots,Y_n}(x_1,\cdots,x_m,y_1,\cdots,y_n;\theta) = c^{m+n}(\theta)\prod_{i=1}^{m}h(x_i)\prod_{j=1}^{n}h(y_j) \cdot \exp\left\{\eta(\theta)\left(\sum_{i=1}^{m}t(x_i)+\sum_{j=1}^{n}t(y_j)\right)\right\}.$$

Letting $T := \sum_{i=1}^{m} t(X_i) + \sum_{j=1}^{n} t(Y_j)$, T is a complete and sufficient statistic for θ, hence the conditional p.m.f. of $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ given T is independent of θ. So, using the conditional distribution of $Y := \sum_{j=1}^{n} t(Y_j)$ given the sufficient statistic T, we can construct a prediction interval which is independent of unknown parameter θ. Actually, we construct a prediction interval of Y according to the following procedures (i) to (iii).

(i) Let $f_{Y|T}(\cdot|t)$ be a conditional p.m.f. of Y given $T = t$. Since T is sufficient for θ, it follows that $f_{Y|T}(\cdot|t)$ is independent of θ. Using $f_{Y|T}(\cdot|t)$, we obtain the conditional mean $\mu_t := E[Y|T = t]$, the conditional variance $\sigma_t^2 := \mathrm{Var}(Y|T = t)$ and the conditional third cumulant $\kappa_{3,t} := \kappa_3(Y|T = t) = E[(Y - \mu_t)^3|T = t]$ of Y given $T = t$.

(ii) Using the Cornish–Fisher expansion with μ_t, σ_t^2 and $\kappa_{3,t}$ in (i), we asymptotically get $\underline{y}(t)$, $\bar{y}(t)$ such that

$$P\{\underline{y}(t) \le Y \le \bar{y}(t)|T = t\} = 1 - \alpha \tag{2}$$

for any α $(0 < \alpha < 1)$ and any $t \in \mathbf{R}^1$.

(iii) From (2), we have for any $\theta \in \Theta$

$$P_\theta\{\underline{y}(T) \le Y \le \bar{y}(T)\} = 1 - \alpha.$$

Since $T := \sum_{i=1}^{m} t(X_i) + \sum_{j=1}^{n} t(Y_j) = \sum_{i=1}^{m} t(X_i) + Y$ is complete and sufficient, we asymptotically obtain $a(\,\cdot\,)$, $b(\,\cdot\,)$ such that

$$P_\theta\{a(\mathbf{X}) \le Y \le b(\mathbf{X})\} = 1 - \alpha.$$

Then the interval $[a(\mathbf{X}),\ b(\mathbf{X})]$ is a prediction interval of Y at confidence coefficient $1-\alpha$.

2.1. Binomial case

Suppose that X is an observable random variable, Y is a random variable to be observed in future, and X and Y are independent. Further, assume that X is distributed according to the binomial distribution $B(m, p)$ whose p.m.f.

$$f_X(x;p) = \binom{m}{x} p^x q^{m-x} \qquad (x = 0, 1, \cdots, m;\ \ 0 < p < 1 \ \text{ and } \ q = 1 - p),$$

and Y is distributed according to the binomial distribution $B(n,p)$, where m and n are known natural numbers, and p is unknown. Then we construct a prediction interval of Y based on X at confidence coefficient $1-\alpha$. Since the joint p.m.f. of (X,Y) is given by

$$f_{X,Y}(x,y;p)=\binom{m}{x}\binom{n}{y}p^{x+y}q^{m+n-(x+y)}$$

$$(x=0,1,\ldots,m;\ y=0,1,\ldots,n;\ 0<p<1,\ q=1-p),$$

it follows that the statistic $T:=X+Y$ is sufficient for p, and T is distributed according to the binomial distribution $B(m+n,p)$. Then the conditional p.m.f. of Y given $T=t$ is

$$f_{Y|T}(y|t)=\frac{\binom{n}{y}\binom{m}{t-y}}{\binom{m+n}{t}}\qquad(\max(0,t-m)\le y\le\min(t,n)),$$

which is independent of p. This means that the prediction interval of Y based on the sufficient statistic T is constructed independently of p. The distribution with the above p.m.f. $f_{Y|T}(y|t)$ is called the hypergeometric distribution $H(t,n,m+n)$. When $T=t$ is given, the conditional mean μ_t, the conditional variance σ_t^2 and the conditional third cumulant $\kappa_{3,t}$ of Y are given by

$$\begin{aligned}\mu_t &= E[Y|T=t]=\frac{tn}{m+n},\\ \sigma_t^2 &= \mathrm{Var}(Y|T=t)=\frac{tmn(m+n-t)}{(m+n)^2(m+n-1)},\\ \kappa_{3,t} &= \kappa_3(Y|T=t)=\frac{tmn(m-n)(m+n-t)(m+n-2t)}{(m+n)^3(m+n-1)(m+n-2)},\end{aligned}$$

respectively.

When m and n are large, using the Cornish–Fisher expansion we asymptotically obtain the upper $100(\alpha/2)$ percentile $y_{\alpha/2}(t)$ of the hypergeometric distribution $H(t,n,m+n)$ such that

$$P\{\min(t,n)-y_{\alpha/2}(t)\le Y\le y_{\alpha/2}(t)|T=t\}=1-\alpha. \tag{3}$$

First, by the Cornish–Fisher expansion we have

$$\frac{y_{\alpha/2}(t)-\mu_t+\frac{1}{2}}{\sigma_t}=u_{\alpha/2}+\frac{\kappa_{3,t}}{6\sigma_t^3}u_{\alpha/2}^2+\cdots,$$

that is,

$$\begin{aligned} y_{\alpha/2}(t) &= \mu_t - \frac{1}{2} + \sigma_t u_{\alpha/2} + \frac{\kappa_{3,t}}{6\sigma_t^2}u_{\alpha/2}^2 + \cdots \\ &= \frac{tn}{m+n} - \frac{1}{2} + u_{\alpha/2}\sqrt{t\left(1-\frac{t}{m+n}\right)\frac{mn}{(m+n)(m+n-1)}} \\ &\quad + \frac{m-n}{6(m+n-2)}\left(1-\frac{2t}{m+n}\right)u_{\alpha/2}^2 + \cdots, \end{aligned} \tag{4}$$

where $u_{\alpha/2}$ is the upper $100(\alpha/2)$ percentile of the standard normal distribution $N(0,1)$. Letting $y := y_{\alpha/2}(t)$, $a := n/(m+n)$, $b := mn/\{(m+n)(m+n-1)\}$, $c := (m-n)/(m+n-2)$, $u = u_{\alpha/2}$ and $t := x+y$, then we obtain from (4)

$$y \fallingdotseq a(x+y) - \frac{1}{2} + u\sqrt{(x+y)\left(1-\frac{x+y}{m+n}\right)b} + \frac{c}{6}\left(1-\frac{2(x+y)}{m+n}\right)u^2, \tag{5}$$

which implies that

$$\left[\left\{1-a+\frac{cu^2}{3(m+n)}\right\}y - \left\{a-\frac{cu^2}{3(m+n)}\right\}x - \frac{c}{6}u^2 + \frac{1}{2}\right]^2 \fallingdotseq b(x+y)\left(1-\frac{x+y}{m+n}\right)u^2.$$

Hence we have

$$\begin{aligned} &\left\{1-a+\frac{cu^2}{3(m+n)}\right\}^2 y^2 + \left\{a-\frac{cu^2}{3(m+n)}\right\}^2 x^2 + \frac{c^2}{36}u^4 + \frac{1}{4} \\ &-2\left\{1-a+\frac{cu^2}{3(m+n)}\right\}\left\{a-\frac{cu^2}{3(m+n)}\right\}xy \\ &+\left\{\frac{c}{3}u^2-1\right\}\left\{a-\frac{cu^2}{3(m+n)}\right\}x - \left\{\frac{c}{3}u^2-1\right\}\left\{1-a+\frac{cu^2}{3(m+n)}\right\}y \\ &-b(x+y)u^2 + \frac{b(x+y)^2}{m+n}u^2 - \frac{c}{6}u^2 = 0, \end{aligned}$$

which implies that

$$\left[\left\{1-a+\frac{cu^2}{3(m+n)}\right\}^2+\frac{bu^2}{m+n}\right]y^2$$

$$-2\left[\left\{1-a+\frac{cu^2}{3(m+n)}\right\}\left\{a-\frac{cu^2}{3(m+n)}\right\}-\frac{bu^2}{m+n}\right]xy$$

$$+\left[\left\{a-\frac{cu^2}{3(m+n)}\right\}^2+\frac{bu^2}{m+n}\right]x^2-\left[\left\{\frac{c}{3}u^2-1\right\}\left\{1-a+\frac{cu^2}{3(m+n)}\right\}+bu^2\right]y$$

$$+\left[\left\{\frac{c}{3}u^2-1\right\}\left\{a-\frac{cu^2}{3(m+n)}\right\}-bu^2\right]x+\frac{c^2}{36}u^4+\frac{1}{4}-\frac{c}{6}u^2=0. \qquad (6)$$

Putting

$$A:=\left\{1-a+\frac{cu^2}{3(m+n)}\right\}^2+\frac{bu^2}{m+n} \quad ,$$

$$B:=\left\{1-a+\frac{cu^2}{3(m+n)}\right\}\left\{a-\frac{cu^2}{3(m+n)}\right\}-\frac{bu^2}{m+n} \quad ,$$

$$C:=\left\{a-\frac{cu^2}{3(m+n)}\right\}^2+\frac{bu^2}{m+n} \quad ,$$

$$2D:=\left\{\frac{c}{3}u^2-1\right\}\left\{1-a+\frac{cu^2}{3(m+n)}\right\}+bu^2,$$

$$2E:=\left\{\frac{c}{3}u^2-1\right\}\left\{a-\frac{cu^2}{3(m+n)}\right\}-bu^2,$$

$$F:=\frac{c^2}{36}u^4+\frac{1}{4}-\frac{c}{6}u^2,$$

we have from (6)

$$Ay^2-2(Bx+D)y+Cx^2+2Ex+F=0$$

whose solution is given by

$$y = \frac{1}{A}\left\{Bx + D \pm \sqrt{(Bx+D)^2 - A(Cx^2 + 2Ex + F)}\right\}.$$

From (3), we asymptotically get a prediction interval $[a(X), b(X)]$ of Y at confidence coefficient $1-\alpha$ such that

$$P_p\{a(X) \le Y \le b(X)\} \fallingdotseq 1 - \alpha$$

for $0 < p < 1$. Then $a(X)$ and $b(X)$ are given by

$$a(X) = \frac{1}{A}\left\{Bx + D - \sqrt{(Bx+D)^2 - A(Cx^2 + 2Ex + F)}\right\},$$

$$b(X) = \frac{1}{A}\left\{Bx + D + \sqrt{(Bx+D)^2 - A(Cx^2 + 2Ex + F)}\right\}.$$

Drawing the curves $Y = a(X)$ and $Y = b(X)$, i.e. the prediction curves of Y, we can get the prediction interval of Y in Figures 2 and 3.

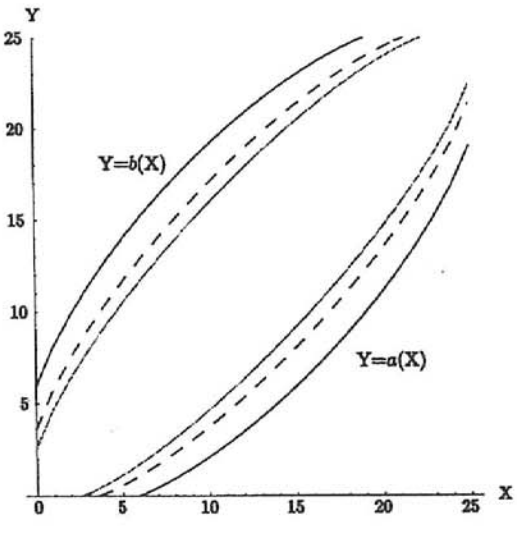

Figure 2: Prediction curves $Y = a(X)$ and $Y = b(X)$ for Y at confidence coefficient $1-\alpha$ for $m = n = 25$

$1-\alpha$: ———— 99% , – – – – 95% , ———— 90%

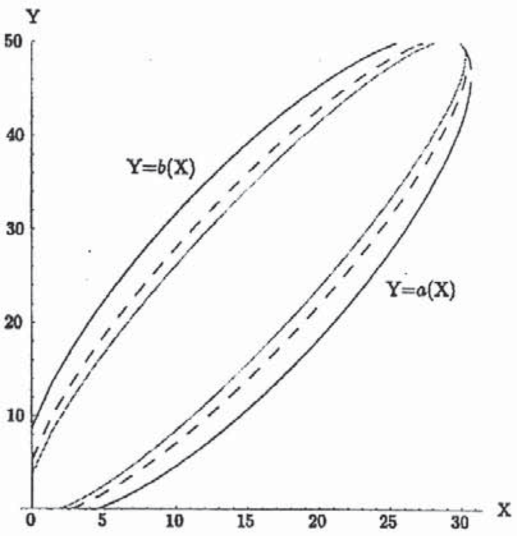

Figure 3: Prediction curves $Y = a(X)$ and $Y = b(X)$ for Y at confidence coefficient $1 - \alpha$ for $m = 30$ and $n = 50$

$1-\alpha$: ———— 99% , — — — — 95% , ———— 90%

2.2. Poisson case

Suppose X is an observable random variable, Y is a random variable to be unobserved, and X and Y are independent. Further, we assume that X is distributed according to the Poisson distribution $Po(m\lambda)$ whose p.m.f.

$$f_X(x) = \frac{e^{-m\lambda}(m\lambda)^x}{x!} \qquad (x = 0, 1, 2, \cdots; \ \lambda > 0),$$

and Y is distributed according to the Poisson $Po(n\lambda)$, when m and n are known natural numbers, and λ is unknown. Then we construct a prediction interval of Y based on X. Since the joint p.m.f. of (X, Y) is given by

$$f_{X,Y}(x, y; \lambda) = \frac{e^{-(m+n)\lambda} m^x n^y \lambda^{x+y}}{x! y!}$$

$$(x = 0, 1, 2, \ldots; \ y = 0, 1, 2, \ldots; \ m, n = 1, 2, \ldots; \ \lambda > 0),$$

it follows that the statistic $T := X + Y$ is sufficient for λ, and T is distributed according to the Poisson distribution $Po((m+n)\lambda)$. Then the conditional p.m.f. of Y given $T = t$ is the binomial distribution $B(t, n/(m+n))$ which is independent of λ. This means that the prediction interval of Y based on the sufficient statistic T is constructed independently of unknown parameter λ. When $T = t$ is given, the conditional mean μ_t, the conditional variance σ_t^2 and the conditional third cumulant $\kappa_{3,t}$ of Y are given by

$$\mu_t = E[Y|T=t] = \frac{tn}{m+n},$$

$$\sigma_t^2 = \mathrm{Var}(Y|T=t) = \frac{tmn}{(m+n)^2},$$

$$\kappa_{3,t} = \kappa_3(Y|T=t) = \frac{tmn(m-n)}{(m+n)^3},$$

respectively. When m and n are very large, in a similar way to (3) using the Cornish–Fisher expansion we asymptotically obtain the upper $100(\alpha/2)$ percentile $y_{\alpha/2}(t)$ of the binomial distribution $B(t, n/(m+n))$ such that

$$P\{t - y_{\alpha/2}(t) \leq Y \leq y_{\alpha/2}(t)|T=t\} = 1-\alpha. \tag{7}$$

By the Cornish–Fisher expansion we have

$$\frac{y_{\alpha/2}(t) - \mu_t + \frac{1}{2}}{\sigma_t} = u_{\alpha/2} + \frac{\kappa_{3,t}}{6\sigma_t^3}u_{\alpha/2}^2 + \cdots,$$

that is,

$$\begin{aligned} y_{\alpha/2}(t) &= \mu_t - \frac{1}{2} + \sigma_t u_{\alpha/2} + \frac{\kappa_{3,t}}{6\sigma_t^2}u_{\alpha/2}^2 + \cdots \\ &= \frac{nt}{m+n} - \frac{1}{2} + u_{\alpha/2}\sqrt{\frac{mnt}{(m+n)^2}} + \frac{m-n}{6(m+n)}u_{\alpha/2}^2 + \cdots, \end{aligned} \tag{8}$$

where $u_{\alpha/2}$ is the upper $100(\alpha/2)$ percentile of the standard normal distribution $N(0,1)$. Letting $y := y_{\alpha/2}(t)$, $a := n/(m+n)$, $b := mn/(m+n)^2$, $c := (m-n)/\{6(m+n)\}$, $u = u_{\alpha/2}$ and $t := x+y$, then we obtain from (8)

$$y \fallingdotseq a(x+y) - \frac{1}{2} + u\sqrt{b(x+y)} + cu^2, \tag{9}$$

which implies that

$$\left\{y - a(x+y) - cu^2 + \frac{1}{2}\right\}^2 \fallingdotseq b(x+y)u^2.$$

Hence we have

$$\begin{aligned} (1-a)^2y^2 &+ 2\left\{(a^2-a)x + acu^2 - cu^2 - \frac{1}{2}bu^2 + \frac{1}{2} - \frac{a}{2}\right\}y \\ &+ a^2x^2 + 2\left(acu^2 - \frac{1}{2}bu^2 - \frac{a}{2}\right)x + c^2u^4 + \frac{1}{4} - cu^2 = 0. \end{aligned} \tag{10}$$

Putting $A := (1-a)^2,\ B := a-a^2,\ C := a^2,\ D := -\{acu^2 - cu^2 - (bu^2/2) - 1/2 + a/2\}$, $E := acu^2 - (bu^2/2) - a/2,\ \ F := c^2u^4 + 1/4 - cu^2$, we have from (10)

$$Ay^2 - 2(Bx + D)y + Cx^2 + 2Ex + F = 0$$

whose solution is given by

$$y = \frac{1}{A}\left\{Bx + D \pm \sqrt{(Bx+D)^2 - A(Cx^2 + 2Ex + F)}\right\}.$$

From (7), we asymptotically get a prediction interval $[a(X), b(X)]$ of Y at confidence coefficient $1-\alpha$ such that

$$P_\lambda\{a(X) \le Y \le b(X)\} \fallingdotseq 1 - \alpha$$

for $\lambda > 0$. Then $a(X)$ and $b(X)$ are given by

$$a(X) = \frac{1}{A}\left\{Bx + D - \sqrt{(Bx+D)^2 - A(Cx^2 + 2Ex + F)}\right\},$$

$$b(X) = \frac{1}{A}\left\{Bx + D + \sqrt{(Bx+D)^2 - A(Cx^2 + 2Ex + F)}\right\}.$$

Drawing the curves $Y = a(X)$ and $Y = b(X)$, i.e. the prediction curves for Y, we can get the prediction interval of Y in Figures 4 and 5.

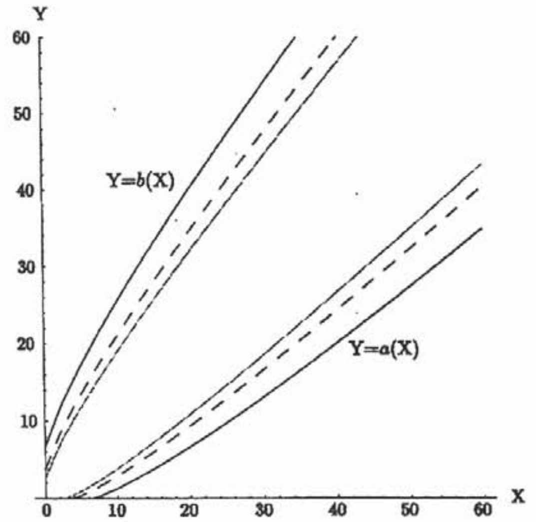

Figure 4: Prediction curves $Y = a(X)$ and $Y = b(X)$ for Y at confidence coefficient $1-\alpha$ for $m = n = 25$

$1-\alpha$: ———— 99% , – – – – 95% , ———— 90%

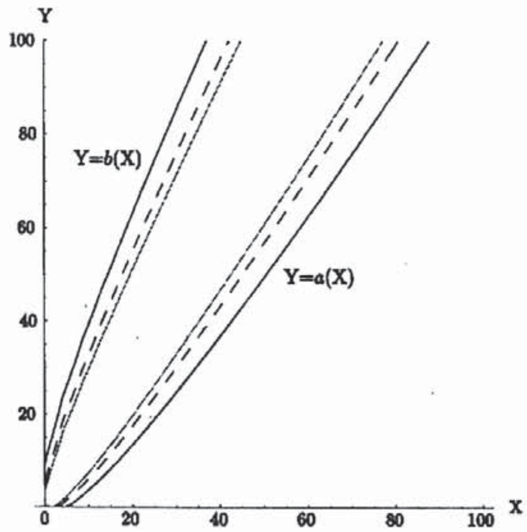

Figure 5: Prediction curves $Y = a(X)$ and $Y = b(X)$ for Y at confidence coefficient $1 - \alpha$ for $m = 30$ and $n = 50$

$1-\alpha$: ———— 99% , – – – – 95% , ———— 90%

2.3. Randomized prediction function

In the previous sections, we consider a non-randomized prediction interval, but we need to take a randomized prediction interval to attain the confidence coefficient $1 - \alpha$ (Takeuchi, 1975).

If for any $\alpha(0 < \alpha < 1)$ there exists an interval $[a(\mathbf{X}), b(\mathbf{X})]$ such that

$$P_\theta\{a(\mathbf{X}) \leq Y \leq b(\mathbf{X})\} \geq 1 - \alpha, \tag{11}$$

for all $\theta \in \Theta$, then the interval is called a prediction interval of Y at confidence coefficient $1 - \alpha$. We also define a randomized prediction function ϕ at confidence coefficient $1 - \alpha$ as

$$\phi(\boldsymbol{x}, y) = \begin{cases} 1 & \text{for} \quad a(\boldsymbol{x}, y) \leq y \leq b(\boldsymbol{x}, y), \\ 0 & \text{for} \quad y < a(\boldsymbol{x}, y), \quad y > b(\boldsymbol{x}, y), \end{cases}$$

where $a(\boldsymbol{x}, y)$ and $b(\boldsymbol{x}, y)$ are functions satisfying

$$E_\theta[\phi(\mathbf{X}, Y)] \geq 1 - \alpha. \tag{12}$$

for all $\theta \in \Theta$. Let $\phi(\boldsymbol{x}, y)$ be a randomized prediction function at confidence level $1 - \alpha$, and $\boldsymbol{x}$ be any fixed. Then there exists $y^*(\boldsymbol{x})$ such that $\phi(\boldsymbol{x}, y)$ is monotone increasing in y for $0 \leq y \leq y^*(\boldsymbol{x})$, monotone decreasing in y for $y^*(\boldsymbol{x}) \leq y$. Then the set $\{y | \phi(\boldsymbol{x}, y) \geq u\}$ also becomes an interval $[c(\boldsymbol{x}, u), d(\boldsymbol{x}, u)]$ for all $u (0 \leq u \leq 1)$ when $\boldsymbol{x}$ is arbitrarily fixed. So, letting U be a uniformly distributed random variable over the interval $[0, 1]$, then

$$P_\theta\{c(\mathbf{X}, U) \le Y \le d(\mathbf{X}, U)\} = E_\theta[\phi(\mathbf{X}, Y)]$$

for all $\theta \in \Theta$ and, if we take ϕ such that

$$E_\theta[\phi(\mathbf{X}, Y)] \equiv 1 - \alpha, \tag{13}$$

then we obtain a similar randomized prediction function ϕ at confidence coefficient $1-\alpha$. We also get a randomized prediction interval

$$\{Y|\phi(\mathbf{X}, Y) \ge U\} = [c(\mathbf{X}, U), d(\mathbf{X}, U)]$$

at confidence coefficient $1-\alpha$, based to $\mathbf{X}$. Since, in a discrete exponential family of distributions with a parameter θ, a complete and sufficient statistic $T = T(\mathbf{X})$ for θ exists, hence a necessary and sufficient condition for (13) to hold is

$$E[\phi(\mathbf{X}, Y)|T] = 1 - \alpha. \tag{14}$$

Now, we consider the binomial case in Section 2.1 as a concrete example. Suppose that X is an observable random variable, Y is a random variable to be observed in future, and X and Y are independent. Further, assume that X is distributed according to the binomial distribution $B(m,p)$ and Y is distributed according to the binomial distribution $B(n,p)$, where m and n are known natural numbers, and p is unknown. The statistic $T := X + Y$ is sufficient for p, and T is distributed according to the binomial distribution $B(m+n,p)$. For each $t = 0, 1, \cdots, m+n$ we take a randomized prediction function $\phi_t(y)$ such that

$$\phi_t(y) = \begin{cases} 0 & \text{for } y < y_0(t), \quad y > y_1(t), \\ \gamma_0(t) & \text{for } y = y_0(t), \\ \gamma_1(t) & \text{for } y = y_1(t), \\ 1 & \text{for } y_0(t) < y < y_1(t), \end{cases}$$

where integers $y_0(t)$, $y_1(t)$ $(0 \le y_0(t) \le y_1(t) \le n)$ and $\gamma_0(t)$, $\gamma_1(t)$ $(0 \le \gamma_0(t) < 1, 0 < \gamma_1(t) \le 1)$ are determined by (14). But, the way of the construction of a randomized prediction function $\phi_t(y)$ is not unique. Here, we choose $y_0(t)$, $y_1(t)$, $\gamma_0(t)$ and $\gamma_1(t)$ such that

$$P\{Y < y_0(t)|T = t\} + (1 - \gamma_0(t))P\{Y = y_0(t)|T = t\} = \frac{\alpha}{2},$$

$$P\{Y > y_1(t)|T = t\} + (1 - \gamma_1(t))P\{Y = y_1(t)|T = t\} = \frac{\alpha}{2}.$$

Indeed, we consider the case when $\alpha = 0.05, 0.10$ for $m = n = 20$. Since, in the case, the conditional joint distribution of Y given $T = t$ is symmetric with respect to m and

n, x and $20 - x$, y and $20 - y$, it is enough to consider only the case $0 \leq t \leq 20$. In the case, $\gamma_0(t) \equiv \gamma_1(t)$ and the values of $y_0(t)$, $y_1(t)$, $\gamma_0(t)$ are given by Tables 1 and 2. From Tables 1 and 2, using a uniformly distributed random number over the interval $[0, 1]$, we obtain a randomized prediction interval

$$\{Y | \phi_{X+Y}(Y) \geq U\} = [c(X, U), d(X, U)]$$

at confidence coefficient $1 - \alpha$. As a result, the difference between the non-randomized prediction interval and the randomized one seems to be small (see Figures 6 and 7). It is also easier to construct a non-randomized prediction interval (curve) in a way in Section 2.1 than to do a randomized prediction one.

t	$y_0(t)$	$y_1(t)$	$\gamma_0(t)$
0	0	0	0.975
1	0	1	0.95
2	0	2	0.8974
3	0	3	0.7833
4	0	4	0.5284
5	1	4	0.9902
6	1	5	0.8155
7	1	6	0.4988
8	2	6	0.9666
9	2	7	0.7183
10	2	8	0.2627
11	3	8	0.8467
12	3	9	0.4721
13	4	9	0.9316
14	4	10	0.5943
15	5	10	0.9886
16	5	11	0.6679
17	5	12	0.0807
18	6	12	0.7079
19	6	13	0.1346
20	7	13	0.7207

Table 1: The values of $y_0(t)$, $y_1(t)$, $\gamma_0(t)$ in the randomized prediction function $\phi_t(y)$ for $\alpha = 0.05$

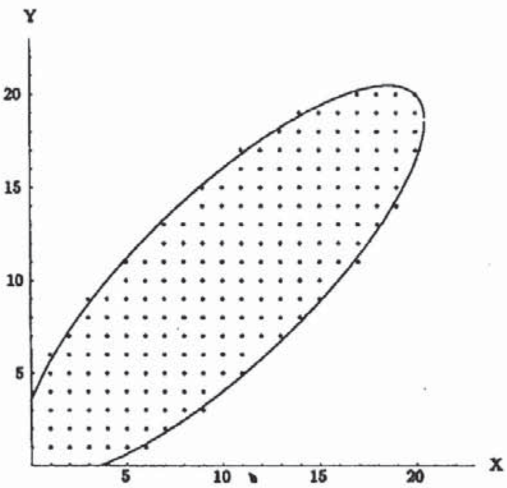

Figure 6: The dots representing the randomized prediction interval of Y based on the randomized prediction function ϕ_t at the confidence coefficient (c.c.) 0.95 and the non-randomized prediction curves at the c.c. 0.95 given in Section 2.1

t	$y_0(t)$	$y_1(t)$	$\gamma_0(t)$
0	0	0	0.95
1	0	1	0.9
2	0	2	0.7947
3	0	3	0.5667
4	0	4	0.0569
5	1	4	0.8206
6	1	5	0.5061
7	2	5	0.9730
8	2	6	0.7055
9	2	7	0.2542
10	3	7	0.8313
11	3	8	0.4442
12	4	8	0.9193
13	4	9	0.5619
14	5	9	0.9815
15	5	10	0.6375
16	5	11	0.0644
17	6	11	0.6835
18	6	12	0.1274
19	7	12	0.7053
20	7	13	0.1472

Table 2: The values of $y_0(t)$, $y_1(t)$, $\gamma_0(t)$ in the randomized prediction function $\phi_t(y)$ for $\alpha = 0.10$

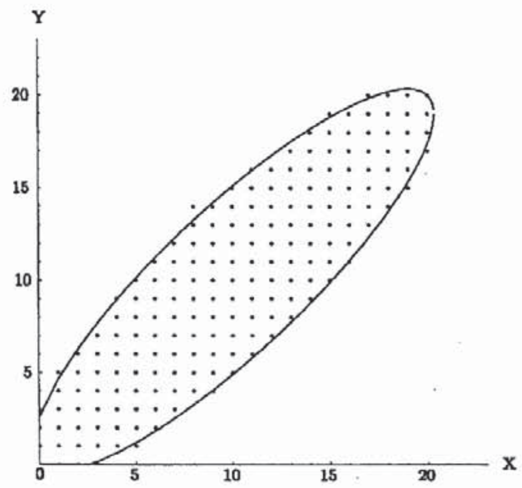

Figure 7: The dots representing the randomized prediction interval of Y based on the randomized prediction function ϕ_t at the confidence coefficient (c.c.) 0.9 and the non-randomized prediction curves at the c.c. 0.9 given in Section 2.1

3. Applications of the prediction interval

First, when some professional baseball team had m games and X wins in them, we consider a prediction interval for the number Y of wins in n residual games, applying to the binomial case. Second, some professional baseball player hited X home runs until certain time, we consider a prediction interval for the number Y of home runs in the rest of games based on X, applying to the Poisson case.

Example 1 (Prediction of the number of wins of the Japanese professional baseball teams). The day, September 10, 1998 was near to the end of the professional baseball season in Japan. In the Central League consisting of six teams, the team "Giants" had the third place but six successive wins up to the day, hence the fans were interested in the final result of the season. So, for the three teams "Bay Stars", "Dragons" and "Giants", we obtain a prediction interval for the number of wins in the rest of games. When each team had m games and X wins in them, we obtain a prediction interval of the number Y of wins in the n games of the rest for the team, applying to the binomial case. Indeed, we get the prediction intervals of Y and prediction curves for Y at confidence coefficient $100(1-\alpha)\%$ including the randomized confidence intervals (see Tables 3 and 4 and Figures 8 to 13).

Team's name	Nos. of finished games	Nos. of wins	Nos. of defeats	No. of draw	Nos. of the rest of games
Bay Stars	110(109)	65	44	1	26
Dragons	115(114)	63	51	1	21
Giants	119	64	55	0	16

Table 3: The result of the three teams in September 10, 1998.

In the above table, (·) means the number of finished games except for the draw. Here, the numbers of the rest of games include those of the draw games, since it is ruled that the draw games are played again in the Central League.

Then we have prediction intervals of the number of wins in the rest of games as follows.

Confidence coefficient(%)	Bay Stars	Dragons	Giants
99	[7.435, 21.748]	[4.483, 17.256]	[2.447, 13.442]
95	[9.227, 20.199]	[6.034, 15.824]	[3.769, 12.197]
90	[10.146, 19.381]	[6.833, 15.074]	[4.452, 11.546]
80	[11.203, 18.419]	[7.757, 14.197]	[5.243, 10.786]
70	[11.914, 17.759]	[8.381, 13.597]	[5.778, 10.267]
60	[12.478, 17.229]	[8.878, 13.117]	[6.203, 9.852]
50	[12.960, 16.770]	[9.303, 12.703]	[6.568, 9.494]
The real numbers of wins in (·) games of the rest	14 (26)	12 (21)	9 (16)

Table 4: The prediction intervals of the number of wins in the rest of games for the three teams "Bay Stars", "Dragons" and "Giants"

Next, at the end of the time of the first half of the season in 1998, that is, in July 21, 1998, the rest of games of the upper three teams was following.

Team's name	Nos. of finished games	Nos. of wins	Nos. of defeats	No. of draw	Nos. of the rest of games
Bay Stars	74	45	28	1	62
Dragons	77	42	34	1	59
Giants	79	41	38	0	56

Table 5: The result of the three teams in July 21, 1998.

Then we obtain prediction intervals of the number of wins at confidence coefficient $100(1-\alpha)\%$ in the latter half of the season in 1998 (see Table 6).

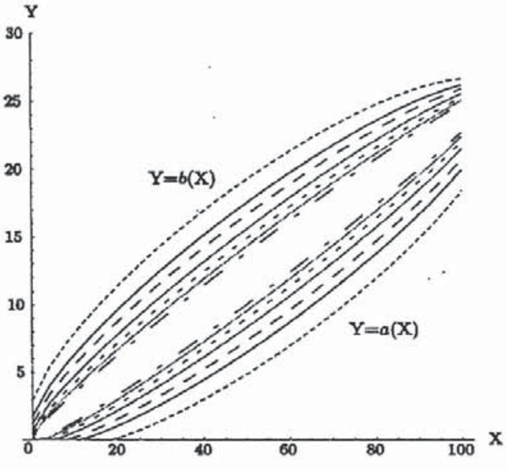

Figure 8: The prediction curves $Y = a(X)$ and $Y = b(X)$ for "Bay Stars"

Confidence coefficient: 99% ; 95%; 90% ; 80% ; 70%; 60% ; 50%

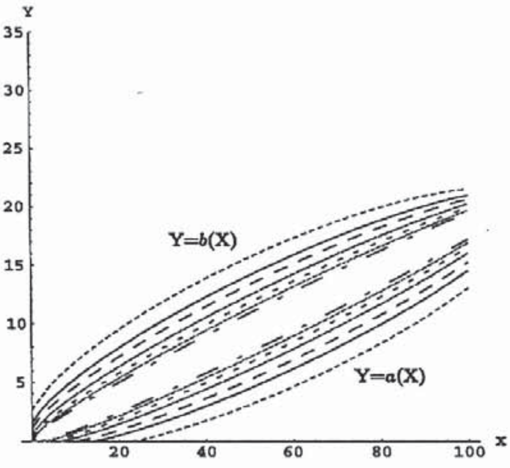

Figure 9: The prediction curves $Y = a(X)$ and $Y = b(X)$ for "Dragons"

Confidence coefficient: 99% ; 95%; 90% ; 80% ; 70%; 60% ; 50%

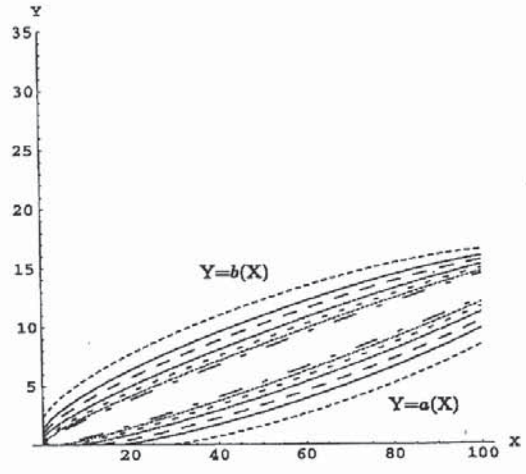

Figure 10: The prediction curves $Y = a(X)$ and $Y = b(X)$ for "Giants"

Confidence coefficient: 99%; 95%; 90%; 80%; 70%; 60%; 50%

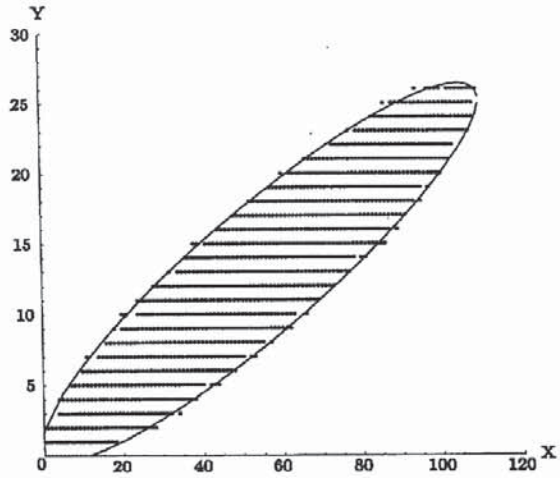

Figure 11: The dots representing the randomized prediction interval for "Bay Stars" based on the randomized prediction function at the confidence coefficient (c.c.) 0.95 and the non-randomized prediction curves at the c.c. 0.95 given Section 2.1

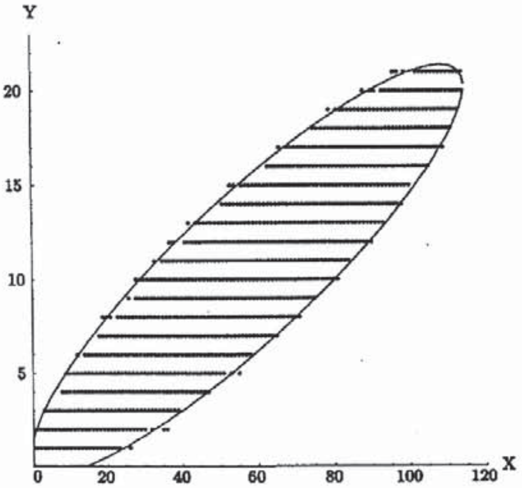

Figure 12: The dots representing the randomized prediction interval for "Dragons" based on the randomized prediction function at the confidence coefficient (c.c.) 0.95 and the non-randomized prediction curves at the c.c. 0.95 given Section 2.1

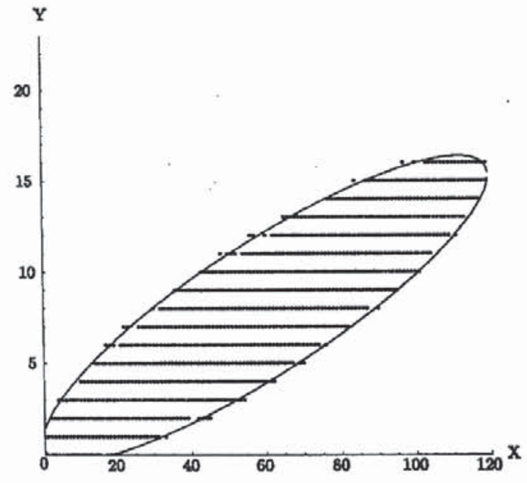

Figure 13: The dots representing the randomized prediction interval for "Giants" based on the randomized prediction function at the confidence coefficient (c.c.) 0.95 and the non-randomized prediction curves at the c.c. 0.95 given Section 2.1

Confidence coefficient(%)	Bay Stars	Dragons	Giants
99	[23.4401, 49.8896]	[18.5622, 44.3707]	[15.7539, 40.5331]
95	[26.7683, 47.0776]	[21.6591, 41.4794]	[18.6568, 37.6838]
90	[28.4767, 45.5848]	[23.2636, 39.9608]	[20.1675, 36.1955]
80	[30.4449, 43.8215]	[25.1249, 38.1810]	[21.9261, 34.4582]
70	[31.7693, 42.6079]	[26.3854, 36.9646]	[23.1207, 33.2751]
60	[32.8184, 41.6306]	[27.3885, 35.99]	[24.0737, 32.3295]
50	[33.7154, 40.7836]	[28.2495, 35.1489]	[24.8932, 31.5152]
The real numbers of wins in the latter half	34	33	32

Table 6: The prediction intervals of the number of wins for the three teams "Bay Stars", "Dragons" and "Giants" in the latter half

We also get the prediction curves of wins of the three teams at confidence coefficient $100(1-\alpha)\%$ in the latter half (see Figures 14 to 16). From the above, we see that the way of construction of a prediction interval in the binomial case in Section 2.1 seems to be reasonable.

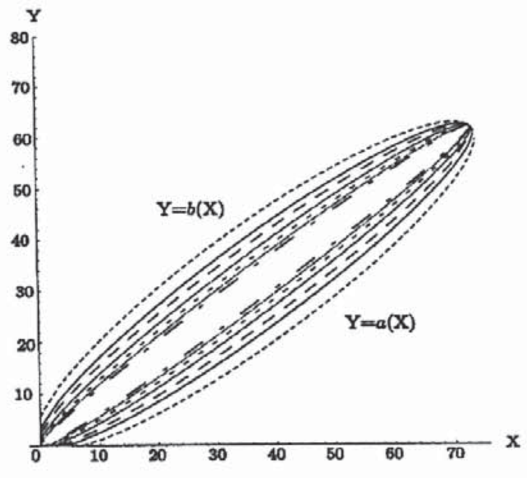

Figure 14: The prediction curves $Y = a(X)$ and $Y = b(X)$ for "Bay Stars"

Confidence coefficient: .... 99% ; ———— 95%; - — — — — 90% ; ———— 80% ; - - - - - - - - 70%; ———— 60% ; —— - —— - 50%

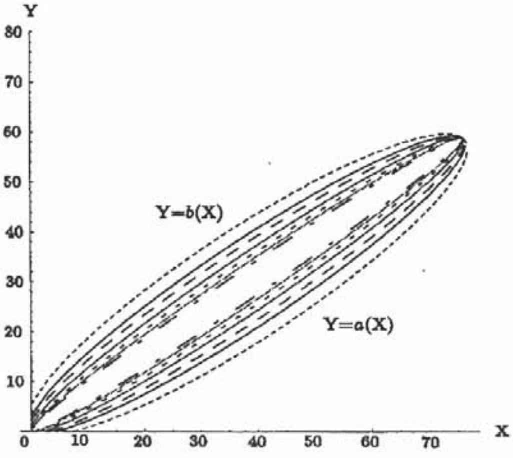

Figure 15: The prediction curves $Y = a(X)$ and $Y = b(X)$ for "Dragons"

Confidence coefficient: 99% ; 95%; 90% 80% ; 70%; 60% 50%

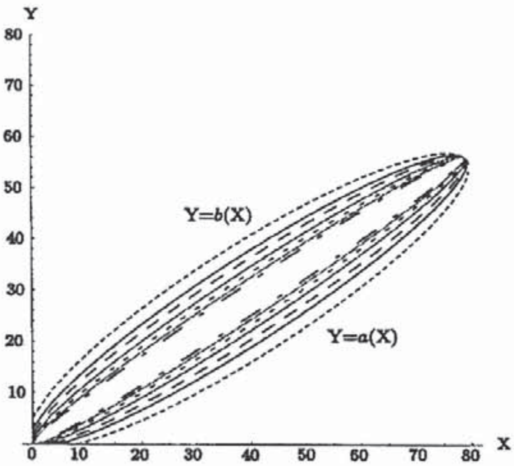

Figure 16: The prediction curves $Y = a(X)$ and $Y = b(X)$ for "Giants"

Confidence coefficient: 99% ; 95%; 90% 80% ; 70%; 60% 50%

Example 2 (Prediction of the number of home runs in the major league). In the major league in the United States, Mark McGwire and Sammy Sosa hited 61 and 58 home runs in September 8, 1998, respectively. When a player hited X home runs in the finished games, we obtain a prediction interval of the number Y of home runs in the rest of games, applying the Poisson case. Indeed, we get the prediction intervals and the prediction curves for Y at confidence coefficient $100(1-\alpha)\%$ including the randomized confidence intervals (see Table 7 and Figures 17 and 18).

Confidence coefficient(%)	McGwire	Sosa
99	[1.071, 16.690]	[0.868, 16.101]
95	[2.371, 14.213]	[2.122, 13.669]
90	[3.091, 13.014]	[2.819, 12.495]
80	[3.968, 11.689]	[3.669, 11.197]
70	[4.589, 10.829]	[4.271, 10.355]
60	[5.099, 10.164]	[4.767, 9.705]
50	[5.549, 9.607]	[5.205, 9.161]
The real number of home runs in 19 games of the rest	9	8

Table 7: The prediction intervals of the number of home runs of McGwire and Sosa in 19 games of the rest

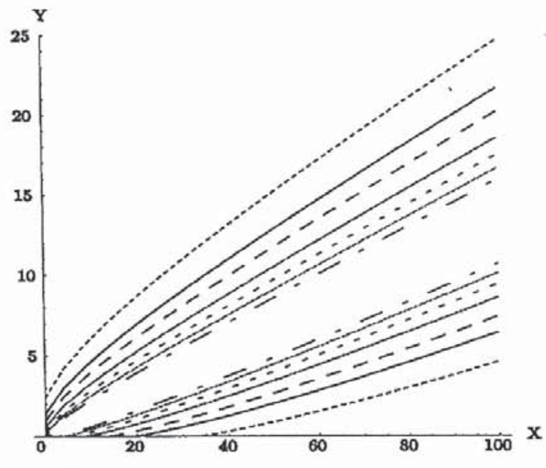

Figure 17: The prediction curves of the number Y of home runs of McGwire and Sosa in 19 games of the rest

Confidence coefficient: ———— 99% ; ———— 95%; - - — — — 90%; ———— 80% ; - - - - - - - - 70%; ———— 60%; — - — - 50%

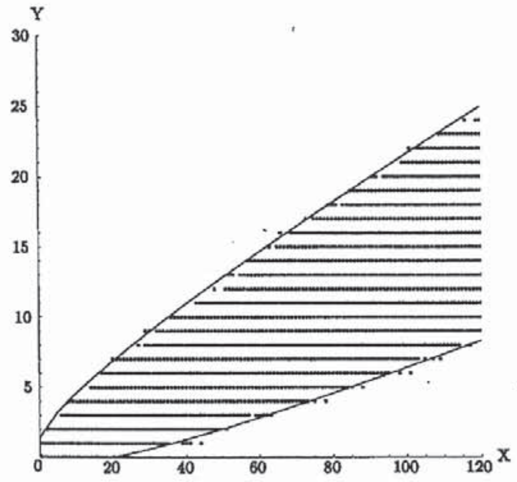

Figure 18: The dots representing the randomized prediction interval for McGwire and Sosa based on the randomized prediction function at the confidence coefficient (c.c.) 0.95 and the non-randomized prediction curves at the c.c. 0.95 given in Section 2.2

Next, at the time when McGwire played 116 games, he hited 46 home runs and the number of his rest of games was 47. On the other hand, at the time when Sosa played 118 games, he hited 44 home runs and the number of his rest of games was 45. Then we get prediction intervals and prediction curves of Y at confidence coefficient $100(1-\alpha)\%$ (see Table 8, Figures 15 to 16).

Confidence coefficient(%)	McGwire	Sosa
99	[6.69146, 33.1830]	[5.57338, 30.5031]
95	[9.05895, 29.1299]	[7.7759 , 26.6592]
90	[10.3445, 27.1589]	[8.97485, 24.7928]
80	[11.8914, 24.9703]	[10.4201, 22.7229]
70	[12.9755, 23.5434]	[11.4344, 21.3747]
60	[13.8607, 22.4374]	[12.2635, 20.3306]
The real number of home runs in (·) games of the rest	24 (47)	22 (45)

Table 8: The prediction intervals of the number of home runs of McGwire and Sosa in games of the rest

From the above, we see that the way of construction of a prediction interval in the Poisson case in Section 2.2 seems to be reasonable.

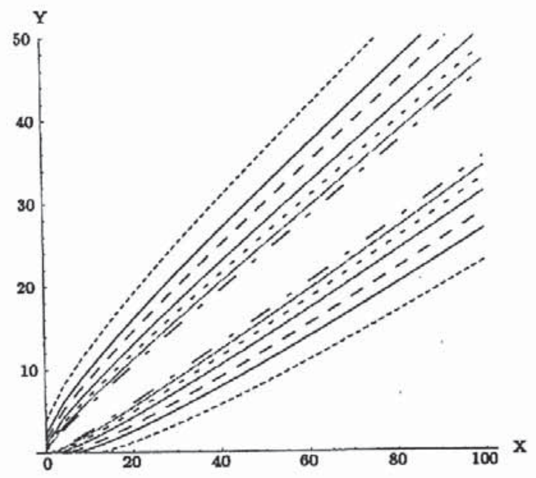

Figure 19: The prediction curves of the number of home runs of McGwire in games of the rest

Confidence coefficient: 99%; 95%; 90%; 80%; 70%; 60%; 50%

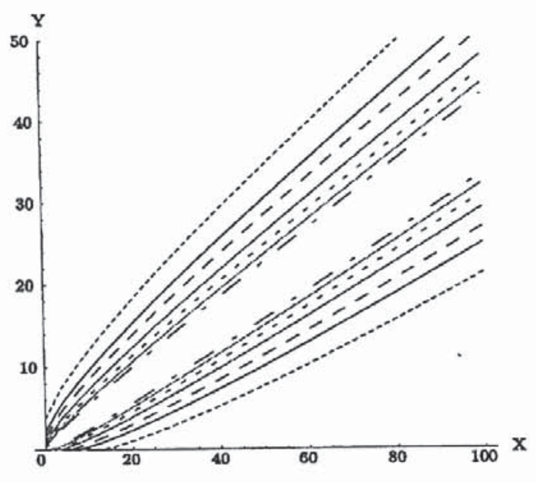

Figure 20: The prediction curves of the number of home runs of Sosa in games of the rest

Confidence coefficient: 99%; 95%; 90%; 80%; 70%; 60%; 50%

REFERENCES

[1] Akahira, M. (1990). *Theory of Statistical Prediction.* Lecture Note at the Middle East Technical University, Ankara.

[2] Barndorff-Nielsen, O. E. and Cox, D. R. (1996). Prediction and asymptotics. *Bernoulli* **2**, 319-340.

[3] Bjφrnstad, J. F. (1990). Predictive likelihood: A review (with discussion). *Statist. Sci.* **5**, 242-265

[4] Butler, R. W. (1986). Predictive likelihood inference with applications (with discussion). *J. R. Statist. Soc.*, **B, 41**,279-312

[5] Geisser, S. (1993). *Predictive Inference: An Introduction.* Chapman & Hall, New York.

[6] Guttman, I. (1970). *Statistical Tolerance Regions: Classical and Bayesian.* Griffin, London.

[7] Hinkley, D. V. (1979). Predictive likelihood. *Ann Statist.* **7**, 718-728

[8] Lauritzen, S. L. (1974). Sufficiency, prediction, and extreme models. *Scand. J. Statist.* **6**. 128-134

[9] Takada, Y. (1996). Statistical properties of prediction intervals. *Sugaku Expositions* **9**, 153-168.

[10] Takeuchi, K. (1975). *Statistical Prediction Theory.* (In Japanese), Baifukan, Tokyo.

ÖZET

X gözlenen rasgele vectör ve Y gelecekde gözlenecek rasgele değişken olsun. **X** ve Y 'nin ortak dağılımının bilinmeyen parametreden bağomlı olduğunu varsayalım. Bu makalede biz kesikli üstel dağılımlar ailesi için Y 'nin **X**'e dayalı öngörü güven aralığını kurmağa çalışıyoruz. Özel halde binomial ve Poisson dağılımları durumunda öngöru güven aralıkları kuruluyor ve pratik problemler üzerinde uygulamalar yapılıyor.

İstatistik, Journal of the Turkish Statistical Association
2000, Vol. 3, No 3, pp. 83-91

A TOTAL EXPANSION OF FUNCTIONAL OF EXIT TIME FROM A SMALL BALL FOR DIFFUSION PROCESS

V.A. Gasanenko
Institute of Mathematics, National Academy of Sciences of Ukraine
Tereshchenkivska 3, 01601, Kiev-4, Ukraine

Abstract

The total expansion on powers $\epsilon \to 0$ of the functional

$$H_\epsilon(\lambda, f) = M[\exp(-\epsilon^{-2}\lambda T_\epsilon)f(\epsilon^{-1}\xi(T_\epsilon))], \lambda \geq 0$$

is obtained. Here $\xi(t)$ is m- dimensional homogeneous diffusion process and T_ϵ is the first exit time of the process $\xi(t)$ from the sphere of radius ϵ. A solution of this problem is reduced to the computation of the total expansion of solutions of some specific elliptic boundary problems.

Key Words: Stochastic equations, small ball, exit time, elliptic boundary problem.

1. Introduction

We consider the solution of m - dimensional stochastic differential equation

$$d\xi(t) = a(\xi(t))dt + \sum_{k=1}^{m} b_k(\xi(t))dw_k(t), \quad \xi(0) = 0.$$

Here $b_k(\cdot), a(\cdot) : R^m \to R^m$ are some smooth functions. Consider the open ball $B_\epsilon = \{x : |x| < \epsilon\}$, and let $B := B_1, S = \partial B_1$. Denote by $T_\epsilon = \inf\{t : \xi(t) \notin B_\epsilon\}$ the first exit time of the process $\xi(t)$ from the sphere $S_\epsilon = \partial B_\epsilon$.

In this paper we deal with the asymptotic expansion of the following functional

$$H_\epsilon(\lambda, f) = M[\exp(-\lambda\frac{T_\epsilon}{\epsilon^2})f(\frac{\xi(T_\epsilon)}{\epsilon})], \quad \lambda \geq 0$$

This problem for Brownian motion was researched in [1]- [3]. In these papers only a few first members of corresponding expansion were defined when $w(t)$ takes values in some manifold. In particular, in [3] the forms for the first three members of expansion were obtained with the help of probability approach.

The physics problems which led to research of diffusion processes in small ball are described in [4,5]. The functional $H_\epsilon(\lambda, f)$ is related to observations over a diffusion process in the sphere S_ϵ at a proper scale.

In this paper the procedure of total expansion is suggested for fixed space R^m. Our approach here is based on the expansion of the solution of related elliptic boundary problems with small parameters.

2. The main results

Let us define the differential operator in the space $C^2(\overline{B})$.

$$Lu(x) = \sum_{i=1}^{m} a_i(x)\frac{\partial u(x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m}\frac{\partial^2 u(x)}{\partial x_i \partial x_j}\sum_{k=1}^{m} b_{ki}(x))b_{kj}(x))$$

As it is known [6], $H_\epsilon(\lambda, f) = v(\epsilon^{-2}\lambda, 0)$ where $v(\epsilon^{-2}\lambda, y)$ is a solution of the following boundary problem

$$\epsilon^{-2}\lambda v(\epsilon^{-2}\lambda, y) - Lv(\epsilon^{-2}\lambda, y) = 0; \quad y \in B_\epsilon, \ v(\epsilon^{-2}\lambda, x) = f(\frac{x}{\epsilon}), \ x \in \partial B_\epsilon.$$

After the change of the variable $z = y/\epsilon$, the function $v_\epsilon(\lambda, z) := v(\epsilon^{-2}\lambda, \epsilon z)$ will be a solution to the following problem:

$$\epsilon^{-2}\lambda v_\epsilon(\lambda, z) - L_\epsilon v_\epsilon(\lambda, z) = 0; \quad z \in B;$$

$$v_\epsilon(\lambda, z) = f(z), \quad z \in S$$

Here

$$L_\epsilon u(z) := \epsilon^{-1}\sum_{i=1}^{m} a_i(\epsilon z)\frac{\partial u(z)}{\partial z_i} + \frac{\epsilon^{-2}}{2}\sum_{i,j=1}^{m}\frac{\partial^2 u(z)}{\partial z_i \partial z_j}\sum_{k=1}^{m} b_{ki}(\epsilon z)b_{kj}(\epsilon z).$$

The expansion of solution v_ϵ we seek in the form

$$v_\epsilon(\lambda; z) = \sum_{k\geq 0}\epsilon^k u_k(\lambda, z).$$

To find the system of equations for the functions v_k, we assume at first that $a_i(x)$ and $b_{ki}(x), 1 \leq k, i \leq m$ are analytical functions in zero. That means the following Taylor-series expansions take place

$$a_i(\epsilon z) = \sum_{l\geq 0} a_{il}(z)\epsilon^l, \quad b_{kj}(\epsilon z) = \sum_{l\geq 0} b_{kjl}(z)\epsilon^l. \tag{1}$$

By definition it means, that

$$a_{il}(z) = \frac{1}{l!} \sum_{1\le j_1 \le \cdots \le j_l \le m} \frac{\partial^l a_i(0)}{\partial z_{j_1} \cdot \cdots \cdot \partial z_{j_l}} z_{j_1} \cdot \cdots \cdot z_{j_l}.$$

$$b_{kjl}(z) = \frac{1}{l!} \sum_{1\le j_1 \le \cdots \le j_l \le m} \frac{\partial^l b_{kj}(0)}{\partial z_{j_1} \cdot \cdots \cdot \partial z_{j_l}} z_{j_1} \cdot \cdots \cdot z_{j_l}.$$

Denote

$$A_{kijl}(z) := \sum_{s=0}^{l} b_{kjs}(z) b_{kil-s}(z), \quad L_0 u(x) = \frac{1}{2} \sum_{1\le i,j \le m} \frac{\partial^2 u(x)}{\partial z_i \partial z_i} \sum_{k=1}^{m} b_{kj}(0) b_{ki}(0).$$

Thus, if $a_i(z)$ and $b_{ki}(z)$ are analytical functions then the operator L_ϵ has the following form:

$$L_\epsilon = \epsilon^{-2} L_0 + \epsilon^{-1} L_1 + \cdots + \epsilon^n L_{n+2} + \cdots,$$

Here

$$L_n = \sum_{i=1}^{m} a_{in-1}(z) \frac{\partial}{\partial z_i} + \frac{1}{2} \sum_{1\le i,j\le m} \frac{\partial^2}{\partial z_i \partial z_j} \sum_{k=1}^{m} A_{kijn}(z), \quad a_{i,-1} = 0.$$

Formally the functions $u_k(\lambda, z), k \ge 0$ should be the solutions of the following system of elliptic problems

$$k = 0: \quad \lambda u_0(\lambda, z) - L_0 u_0(\lambda, z) = 0, z \in \text{int} B; \quad u_0(\lambda, z) = f(z), z \in S,$$

$$k \ge 1: \lambda u_k(\lambda, z) - L_0 u_k(\lambda, z) = \sum_{s=1}^{k} L_s u_{k-s}, z \in \text{int} B; \quad u_k(\lambda, z) = 0, z \in S \qquad (2)$$

Thus, free members of the operator part of the k-th problem $k \ge 1$ in the system (2) are defined with the help of first k coefficients from the expansions (1) and solutions of the previous problems. For further analysis of the system (2) we shall introduce some definitions and results from the theory of elliptic equations [5].

Definitions:

Ω - bounded area in m - dimensional Euclidean space E_m,

S - boundary of Ω, $\overline{\Omega}$ - completion of Ω,

$osc\{u(x); \Omega\}$- fluctuation $u(x)$ on Ω:

$$osc\{u(x); \Omega\} = vrai \max_{\Omega} u(x) - vrai \min_{\Omega} u(x),$$

K_ρ is an arbitrary open ball in space E_m, $\quad \Omega_\rho := K_\rho \cap \Omega$.

We consider that a function $u(x)$ satisfies Holder condition with parameter α; $\alpha \in (0, 1)$ and bounded Holder constant $|u|_{(\alpha),\Omega}$ in domain $\overline{\Omega}$ if the following equality takes place

$$\sup \rho^{-\alpha} osc\{u; \Omega_\rho^i\} = |u|_{(\alpha),\Omega},$$

where sup is calculated over all connected components Ω_ρ^i of all $\Omega_\rho, \rho \le \rho_0$.

Let $C_{0,\alpha}(\Omega)$ be a Banach space of all continuous functions $u(x)$ $x \in \Omega$, with the bounded norm $|u|_{(\alpha),\Omega}$. The norm in $C_{0,\alpha}$ is defined as follows

$$|u|_{\alpha,\Omega} = \max_{\Omega} |u| + |u|_{(\alpha),\Omega}.$$

Here, D^k is a symbol of the derivative $u(x)$ in x of the k order, $C_{l,\alpha}(\overline{\Omega})$ is Banach space of continuous in Ω functions with continuous derivative in Ω of the l order, and the following value is bounded

$$|u|_{l,\alpha,\Omega} = \sum_{k=0}^{l} \sum_{(k)} \max_{\Omega} |D^k(x))| + \sum_{(l)} |D^l u|_{\alpha,\Omega}.$$

Here, symbol $\sum\limits_{(k)}$ denotes the summation over all derivatives of order k.

It is possible to consider elements from $C_{l,\alpha}(\overline{\Omega})$ as functions which are continuous and l times continuously differentiable in $\overline{\Omega}$. To do this it is necessary to complete a definition of $u(x)$ and its derivatives on boundary S by a continuity.

Introduce the classification of boundaries. Let $x^0 = (x_1^0, \cdots, x_m^0)$ be some point of boundary S of domain Ω. We consider that $(y_1, \cdots, y_m)$ is a local cartesian system of coordinates with the centre in a point x^0, if y and x are connected by equality $y_i = d_{ik}(x_k - x_k^0), i = 1, \cdots m$, where d_{ik} is the orthogonal numerical matrix and the axis y_n is directed along external normal (in respect to Ω) to S in a point x^0.

We say, a surface S belongs to class $C_{l,\alpha}, l \ge 1, \alpha \in [0,1]$, if there is a value $\rho > 0$ such, that the intersection of S with a sphere K_ρ having the centre at a point $x^0 \in S$ is a connected surface. The equation of this surface in a local cartesian system of axes $(y_1, \dots, y_n)$ with the centre in a point x^0 has the form $y_n = \omega(y_1, \cdots, y_{n-1})$, and $\omega(y_1, \cdots, y_{n-1})$ is a function of the class $C_{l,\alpha}$ in the domain $\overline{\Omega}$, and this function being projection of $K_\rho \cap S$ on plane $y_n = 0$.

Let us consider the equation

$$Au := \sum_{1 \le i,j \le m} a_{i,j}(x) u_{x_i x_j} + \sum_i a_i(x) u_{x_i} + a(x) u = f(x). \tag{3}$$

Coefficients of this equation and free member $f(x)$ are defined in the bounded domain Ω and belong to space $C_{l-2,\alpha}(\overline{\Omega}), l \ge 2, \alpha \in (0,1)$.

Assume also that condition $a_{ij} = a_{ji}$ is satisfied and equation (3) is elliptic in $\overline{\Omega}$:

$$\sum_{i,j} a_{i,j}(x) \xi_i \xi_j \ge \nu \sum_{i=1}^{m} \xi_i^2, \quad \nu = const > 0 \qquad (4).$$

For the function $u(x)$ satisfying the equation (3) and the following condition on the boundary S:

$$u|_S = \varphi(s), \tag{5}$$

the following Theorem of existence [5, page.142] takes place.

Theorem 1 If coefficients of A belong to the space $C_{0,\alpha}(\overline{\Omega})$, satisfy inequalities (4) and $a(x) \leq 0$; S belongs to $C_{2,\alpha}$, then the problem (3), (5) has the unique solution in $C_{2,\alpha}(\overline{\Omega})$ for all $f(x)$ from $C_{0,\alpha}(\overline{\Omega})$ and all $\varphi(s)$ from $C_{2,\alpha}(S)$.

Using Theorem 1 we shall prove, that the partial sum $\sum\limits_{k=0}^{n} u_k\epsilon^k$ is an approximation to v_ϵ.

Theorems 2. Suppose the following conditions are satisfied

1^0. Functions $a_i(x)$ and $b_{ki}(x)$ $1 \leq k, i \leq m$ have n-th and $n+1$- th continuous derivatives in zero.

2^0. $f(z) \in C_{2,\alpha}(S)$.

3^0. The functions $u_k(\lambda, z)$ are the solutions of system (2) if $k \leq n$.

Then functions $u_k(\lambda, z), 0 \leq k \leq n$ and $\mu_\epsilon(\lambda, z) = v_\epsilon(\lambda, z) - \sum\limits_{k=0}^{n} u_k(\lambda, z)\epsilon^k$ satisfy the following relations

$$u_k(\lambda, z) \in C_{2,\alpha}(\bar{B}), \qquad \sup_{\lambda>0,\ z\in\bar{B}} |\mu_\epsilon(\lambda, z)| \leq \epsilon^{n+1}K,$$

where K is a constant.

Proof. From the condition 1^0 the following representations take place

$$a_i(\epsilon z) = \sum_{l=0}^{n-1} a_{il}(z)\epsilon^l + a_{in}(z)\epsilon^n,$$

$$b_{kj}(\epsilon z) = \sum_{l=0}^{n} b_{kjl}(z)\epsilon^l + b_{kjn+1}(z)\epsilon^{n+1}, \quad i, k, j = \overline{1, m},$$

More precisely, the functions $a_{in}(z)$ and $b_{kjn+1}(z)$ have the following form:

$$a_{in}(z) = \frac{1}{n!} \sum_{1\leq j_l\leq m} \frac{\partial^n a_i(\theta(z))}{\partial z_{j_1} \cdots \partial z_{j_n}} z_{j_1} \cdots z_{j_n};$$

$$\theta(z) = (\theta_1\epsilon z_1, \cdots, \theta_m\epsilon z_m), \quad 0 \leq \theta_i \leq 1, \quad i = \overline{1, m}.$$

$$b_{kjn+1}(z) = \frac{1}{(n+1)!} \sum_{1\leq j_l\leq m} \frac{\partial^{n+1} b_{kj}(\beta(z))}{\partial z_{j_1} \cdots \partial z_{j_{n+1}}} z_{j_1} \cdots z_{j_{n+1}};$$

$$\beta(z) = (\beta_1\epsilon z_1, \cdots, \beta_m\epsilon z_m), \quad 0 \leq \beta_i \leq 1, \quad i = \overline{1, m}.$$

Further, conditions of Theorem 2 and Theorem 1 imply that there is a unique solution $u_0(\lambda, z)$ for the first problem in the system (2) such that $u_0 \in C_{2,\alpha}(\bar{B})$. Taking into account this inclusion, the condition 1^0 and the construction of the free member of second problems in (2), we conclude that the free member of this problem belongs to the space $C_{0,\alpha}(\bar{B})$. Now again using Theorem 1, we obtain the unique solution $u_1(\lambda, z)$ of the second problem in the system (2) such that $u_1 \in C_{2,\alpha}(\bar{B})$. By analogy, it can be proved that there exists unique solutions $u_2, \cdots, u_n$ of corresponding problems of the system (2) and $u_i \in C_{2,\alpha}(\bar{B}) \quad i \le n$.

Thus, the function $\mu_\epsilon(\lambda, z)$ is defined correctly. ¿From the definition of functions v_ϵ and $u_k, k = 0, 1, \ldots, n$ and the decomposition of the operator L_ϵ it follows, that μ_ϵ is a solution of the following problem

$$\epsilon^2 L_\epsilon \mu_\epsilon - \lambda \mu_\epsilon =$$

$$= \Bigg(\epsilon^2(\epsilon^{-2} L_0 + \cdots + \epsilon^{n-2} L_n + \epsilon^{n-1} L_{n+1}) +$$

$$+ \sum_{l=n+2}^{2n+2} \epsilon^l \Big(\sum_{s=0}^{l} \sum_{k=1}^{m} \sum_{i,j=1}^{m} \tilde{b}_{kjs}(z) \tilde{b}_{kil-s}(z) \frac{\partial^2}{\partial z_i \partial z_j} \Big) \Bigg) \mu_\epsilon - \lambda \mu_\epsilon =$$

$$= \epsilon^{n+1} \sum_{i=1}^{m} \sum_{l=0}^{n} a_{il}(z) \frac{\partial}{\partial z_i} u_{n-l}(\lambda, z) + \sum_{l=n+1}^{2n} \epsilon^l \Big(\sum_{i=1}^{m} \sum_{s=0}^{l} \tilde{a}_{il}(z) \frac{\partial}{\partial z_i} \tilde{u}_{l-s}(\lambda, z) \Big) +$$

$$+ \sum_{l=n+1}^{3n+2} \epsilon^l \Big(\sum_{k=1}^{m} \sum_{i,j=1}^{m} \sum_{s=0}^{l} \big(\sum_{r=0}^{s} \tilde{b}_{kjr}(z) \tilde{b}_{kis-r}(z) \big) \frac{\partial^2}{\partial z_i \partial z_j} \tilde{u}_{l-s}(\lambda, z) \Big) := \epsilon^{n+1} K_{\epsilon,n}(\lambda, z);$$

here for $1 \le i, j \le m$

$$\tilde{a}_{il}(z) = \begin{cases} a_{il}(z) & ; \text{ if } \ l \le n \\ 0 & ; \text{ if } \ l > n. \end{cases} \quad \tilde{b}_{kjl}(z) = \begin{cases} b_{kjl}(z) & ; \text{ if } \ l \le n+1 \\ 0 & ; \text{ if } \ l > n+1. \end{cases}$$

and by analogy, functions $\tilde{u}_l(\lambda, z)$ are defined as follows:

$$\tilde{u}_l(\lambda, z) = \begin{cases} u_l(z) & ; \text{ if } \ l \le n \\ 0 & ; \text{ if } \ l > n. \end{cases}$$

Here $K_\epsilon(\lambda, z) \in C_{0,\alpha}(\bar{B})$ for any ϵ and $\sup\limits_{\epsilon \ge 0, \ \lambda \ge 0, \ z \in \bar{B}} |K_\epsilon(\lambda, z)| \le K < \infty$.

Following results of the paper [7, page.135], we obtain the following a priori estimation of the function μ_ϵ:

$$\max_{B, \ \lambda \ge 0} |\mu_\epsilon(\lambda, z)| \le \max_{B, \ \lambda} (\alpha - e^{-\beta x_1}) \times$$

$$\times \max \{ \max_{S} | \frac{\mu_\epsilon(\lambda, z)}{\alpha - e^{-\beta x_1}} |, \epsilon^{n+1} \max_{B, \ \lambda \ge 0} \frac{K}{\beta e^{-\beta x_1} [\beta A_{11}(\epsilon x) - a_1(\epsilon x)] + \lambda(\alpha - e^{-\beta x_1})} \}.$$

Here $A_{11}(\epsilon x) := \sum_{k=1}^{m} b_{k1}^2(\epsilon x)$. The constants β and α are fixed positive values, defined from the condition of positiveness of the following functions in the ball $\bar{B}$:

$$\beta e^{-\beta x_1}(A_{11}(\epsilon x)\beta - a_1(\epsilon x)), \quad \alpha - e^{-\beta x_1}.$$

¿From the latter, the following estimation implies

$$\max_{\bar{B}} |\mu_\epsilon(\lambda, z)| \leq \epsilon^{n+1} K_1, \quad K_1 < \infty.$$

The proof is complete.

Thus, under conditions of this Theorem, the following representation is valid:

$$H_\epsilon(\lambda, f) = \sum_{k=o}^{n} \epsilon^k u_k(\lambda, 0) + O(\epsilon^{n+1}).$$

Now we describe the one of possible schemes of the solution of the problems in the system (2).

The solution $u_0(\lambda, z)$ of the first task in (2) we seek as the sum of two functions $u_{01}(z)$ and $u_{02}(\lambda, z)$: $u_0 = u_{01} + u_{02}$. The function u_{01} is the solution of known Dirichlet problem of the elliptic equation

$$L_0 u_{01} = 0; \quad u_{01}(z)|_{z \in S} = f(z).$$

The function u_{02} is the solution of the following nonhomogeneous problem with zero boundary condition

$$L_0 u_{02} = \lambda u_{02} + \lambda u_{01}; \quad u_{02}|_{z \in S} = 0. \tag{3}$$

The function u_{02} we shall define by the Fourier method. Denote the subspace of the functions of $C_{2,\alpha}(\bar{B})$, which accept zero value onto the surface S, by the $H_{0,2,\alpha}(\bar{B})$. Define scalar product in $H_{0,2,\alpha}(\bar{B})$ similar to the paper [8,sec. 3]:

$$(f, g)_{H_{0,k,\alpha}} = \sum_{|s| \leq 2} \int_B D^s f D^s g dx, \quad ||f||_{H_{0,2,\alpha}(B)} = \sqrt{\sum_{|s| \leq 2} \int_{\bar{B}} |D^s f|^2 dx}.$$

Again, similar to [6, sec.3,] it is possible to show, that $H_{0,2,\alpha}(\bar{B})$ is Hilbert space. Let λ_k and $\beta_k(z)$, $\quad k \geq 1$ be corresponding eigenvalues and eigenfunctions of the following task:

$$L_0 \beta(z) = \lambda \beta(z), \quad z \in intB; \qquad \beta(z)|_{z \in S} = 0.$$

As it is known [8, page 191], these eigenvalues are real values and $\lambda_k \to -\infty$ as $k \to \infty$, eigenfunctions form complete orthonormalized system in the space $H_{0,\tilde{2},\alpha}(B)$.

As $u_{02} \in H_{0,2,\alpha}$, the following representation takes place

$$u_{02}(\lambda, z) = \sum_{k\geq 1} c_{0k}(\lambda)\beta_k(z).$$

Further, substituting this relation into (3) after integration we obtain the equations for coefficients $c_{0k}(\lambda)$:

$$\lambda_k c_{0k}(\lambda) = \lambda c_{0k}(\lambda) + \lambda \int_B u_{01}(z)\beta_k(z)dz.$$

This relation implies the following equality

$$c_{0k}(\lambda) = \frac{\lambda}{\lambda_k - \lambda} \int_B u_{01}(z)\beta_k(z)dz.$$

All further solutions have similar representations:

$$u_m(\lambda, z) = \sum_{k\geq 1} c_{mk}(\lambda)\beta_k(z), \quad \text{where,}$$

$$c_{mk}(\lambda) = \frac{\lambda}{\lambda_k - \lambda} \int_B \left(\sum_{s=1}^{m} L_s(z) u_{m-s}(\lambda, z)\right) \beta_k(z)dz.$$

Acknowledgment

The author thanks Professor Vladimir Anisimov for many helpful comments.

References

[1] Pinsky M.A. Local stochastic differential geometry or what can you learn about a manifold by watching Brownian motion? // Contemp. Math. (1988) 73, 263 - 272.

[2] Liao M. Hitting distributions of small geodesic spheres. // Ann. Probab. (1988) 16, 1039–1050.

[3] Hara K. Wiener functionals associated with joint distributions of exit time and position from small geodesic balls. // Ann. Probab. (1996) v.2, 4, 825-837.

[4] Graham R. Path integral formulation of general diffusion processes. // Z. Phys. (1979) B 26, 281–290.

[5] Onsager L. and Machlup S. Fluctuation and irreversible processes. I,II // Phys. Rev. (1953) 91, 1505–1512.

[6] Gikhman, I. I. and Skorokhod, A. V. Introduction to the theory of random processes (1969) W. B. Saunders Company, Philadelphia.

[7] Ladyjenskai O.J., Ural'ceva N.N. Linear and quasilinear equation of elliptic type. Nauka (1964) 540 p. (in Russian)

[8] Mihailov V.P. The differential equations in partial derivateves. Nauka (1983) 424 p. (in Russian)

ÖZET

$H_\epsilon(\lambda, f) = M[\exp(-\epsilon^{-2}\lambda T_\epsilon) f(\epsilon^{-1}\xi(T_\epsilon))]$, $\lambda \geq 0$ fonksiyonelinin ϵ kuvvetinin $\epsilon \to 0$ iken toplam açılımını verdik. Burada $\epsilon(t)$ $m-$ boyutlu homojen diffuzyon süreci ve T_ϵ, $\xi(t)$ süreci için ϵ yarıçaplı küreden ilk çıkış zamanını göstermektedir. Problemin çözümü bazı özel elliptik sınır problemlerinin çözümlerinin toplam açılımının hesaplanmasına indirgenmiştir.

İstatistik, Journal of the Turkish Statistical Association
2000, Vol. 3, No 3, pp. 92-98

RELIABILITY OF A RESTORED SYSTEM FOR A REPAIR DEVICE WITH INSTANTANEOUS SERVICE

Oleksiy V.Yevdokimov
G.Skovoroda State Pedagogical University, Kharkov, Ukraine

Alexander V. Makarychev
Automobile Technical University, Kharkov, Ukraine

Abstract

Let q be probability of failure of a system in a regeneration period. Let τ_0 be time from the moment when all units of a system are in running order to failure of a system. These characteristics of reliability have been investigated. Asymptotic estimates for q and τ_0 are given.

Key Words: failure of a system, probability, distribution function, random process

1. Introduction

Special reliability problems arise in connection with servicing of automatic machines. As early as 1933 A.Y.Khintchine investigated the problem of servicing of some automatic machines by a single repairman [1]. Later the similar problem was studied by C.Palm [2]. In 1954 J.Taylor and R.R.P.Jackson considered the problem of provision of spare aircraft machines [3]. Since then many papers have been published dealing with reliability of different systems with repairing of failing units(for more information we refer to [4], [5], [6] and [7]). A.D.Solovjev studied characteristics of reliability for a general restored system. He obtained the following two-sided estimates for probability q of failure of a system in a regeneration period in the case of one repair device without instantaneous service:

$$b_{n-1}^{(\upsilon)} \le q \le \frac{b_{n-1}^{(\upsilon)}}{1-2^{n-1}b_0^{(\upsilon)}} \quad , \tag{1}$$

where

$$b_{k-1}^{(\upsilon)} = \int_0^\infty \lambda \frac{(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x} P(\upsilon > x) dx, \quad k = 1,2,\mathrm{K}\ ,n \tag{2}$$

are probabilities that at least k units come to the repair device during the time υ of repairing of the first unit of a system [8]. Also, A.D.Solovjev proved the following limiting theorem for regenerative

processes of special type. (A regeneration period consists of two parts, times of each part are independent random variables. The first part ξ has exponential distribution with parameter λ, the second part η has a general distribution with the mean T_0) [8].

Theorem. If $\lambda T_0 \to 0$, then

$$P(\lambda q \tau > x) \to e^{-x}, \tag{3}$$

where τ is time from start of working of a system to failure of it.

2. Statement of Results

Following A.D.Solovjev [8] we shall treat the case when units of a restored system are serviced by a single repair device. However, it has instantaneous service.

The model describing the situation. There are n+1 units in a system. At first one unit is a working one and n units are in reserve. At the moment of failure of the working unit a unit of the reserve takes it's place. The failing unit is headed for the repair device. After repairing it returns in the system and takes a free place in the reserve. As time goes on the repair device requires instantaneous service, which interrupts repairing of a failing unit. After ending of instantaneous service for the repair device this failing unit is repaired in it. Failing units are repaired one by one in the order of coming to the repair device. Repairing for failing units takes place only for lack of instantaneous service for the repair device. Failure of the system occurs if a working unit fails and the reserve is empty at this moment of time.

RELIABILITY OF A RESTORED SYSTEM

Notations and conceptions. Let λ_0 and λ be parameters of exponential distribution for time of lack of instantaneous service for the repair device and for time of trouble-free operation of the units correspondingly.

Let $\Gamma(x)$ and $G(x)$ be distribution functions for time of instantaneous service for the repair device and for time of repairing for a unit correspondingly. Let τ_0 be time from the moment when all units of the system are in running order to failure of the system. Denote by q probability of failure of a system in a regeneration period.

Denote by $\Gamma_k(x)$ the distribution function of the sum of k independent random variables, which are distributed as $\Gamma(x)$.

Times of trouble-free operation of the units, times of repairing for the units and times of lack of instantaneous service for the repair device are independent random variables. Let $\nu(t)$ be a number of failing units of the system at the moment of time t. Random process $\nu(t)$ is a regenerative one. Moments of transition of this process to the state $\{0\}$ (when $\nu(t)=0$) are the moments of regeneration.

Let w_1 be stationary time of possible waiting for start of repairing of the first unit which comes to the repair device in a regeneration period. Let $\gamma_1=\int_0^\infty x d\Gamma(x)$ be mathematical expectation of time of instantaneous service for the repair device.

Denote by p_1 stationary probability of instantaneous service for the repair device and let $p_0=1-p_1$ be probability of the opposite event.

From the theory of alternating processes of regeneration [9],

$$p_0=\frac{1}{1+\lambda_0\gamma_1},\ p_1=\frac{\lambda_0\gamma_1}{1+\lambda_0\gamma_1},$$

$$w_1=\begin{cases}0 & \text{with probability } p_0\\ \theta & \text{with probability } p_1\end{cases},\qquad \text{where } P(\theta>x)=\frac{1}{\gamma_1}\int_x^\infty(1-\Gamma(t))dt.$$

Let υ' be time from the start of repairing of a unit, which comes to the repair device, to the ending of it's repairing, including instantaneous service for the repair device.

Denote by $G_0(x)$ the distribution function for time υ'. It can be found as:

$$G_0(x)=P(\upsilon'\le x)=\int_0^x\sum_{k=0}^\infty\frac{(\lambda_0 t)^k}{k!}e^{-\lambda_0 t}\Gamma_k(x-t)dG(t) \qquad (4)$$

Let $\upsilon = w_1 + \upsilon'$ be time from the moment of failure of the first unit to it's full repairing in a regeneration period. Denote by $b_{k-1}^{(\upsilon)}, k = 1,2,\ldots, n$ the same probabilities as (2).

We improved the two-sided estimates (1) for probability q in the case of a restored system for a repair device with instantaneous service. For these estimates we used another way of proof than A.D.Solovjev did for (1). Using the improved analogue of (1) and the limiting theorem (3) we can find asymptotic estimates for q and τ_0.

The two-sided inequality and asymptotic estimates can be successfully applied in practical situations.

3. Proofs of Results

Lemma 1. For probability q the following inequality is true:

$$b_{n-1}^{(\upsilon)} \cdot C(\lambda_0, \lambda, \gamma_1) \le q \le \frac{b_{n-1}^{(\upsilon)} \cdot C(\lambda_0, \lambda, \gamma_1)}{1 - b_0^{(\upsilon)} \cdot \left(2^{n-1} - 1\right)} \qquad (5),$$

where $C(\lambda_0, \lambda, \gamma_1)$ is a constant which depends on $\lambda_0, \lambda, \gamma_1$.

Proof. Let q_i be probability of failure of the system in an employment period for the repair device, provided that there are i units with full time of repairing in it at the start of this employment period and instantaneous service for the repair device does not happen. Denote by $a_i^{(\upsilon)} = b_{i-1}^{(\upsilon)} - b_i^{(\upsilon)}$ probabilities that i units come to the repair device during the time υ of repairing of the first unit of the system. Then

$$q \le \left(b_{n-1}^{(\upsilon)} + \sum_{i=1}^{n-1} a_i^{(\upsilon)} q_i \right) \cdot C(\lambda_0, \lambda, \gamma_1),$$

$$q_k = b_{n-k}^{(\upsilon')} + \sum_{i=0}^{n-k} a_i^{(\upsilon')} q_{i+k-1}, \quad k = 1,2,\ldots, n-1, \quad (q_0 = 0).$$

Using equalities for $a_i^{(\upsilon)}$ and $a_i^{(\upsilon')}$, we obtain

$$q \le \left(b_{n-1}^{(\upsilon)} + \sum_{i=1}^{n-1} (q_i - q_{i-1}) b_{i-1}^{(\upsilon)} - q_{n-1} b_{n-1}^{(\upsilon)} \right) \cdot C(\lambda_0, \lambda, \gamma_1), \qquad (6)$$

$$q_k - q_{k-1} = b_{n-k}^{(\upsilon')} + \sum_{i=1}^{n-k} (q_{i+k-1} - q_{i+k-2}) b_{i-1}^{(\upsilon')} - q_{n-1} b_{n-k}^{(\upsilon')}.$$

Setting $x_1 = q, x_k = q_k - q_{k-1}$, $k = 2, \ldots, n-1$ and using the following inequalities

$$a_i^{(\upsilon)} \le b_{i-1}^{(\upsilon)},\ a_i^{(\upsilon')} \le b_{i-1}^{(\upsilon')} \le b_{i-1}^{(\upsilon)},$$

we obtain such system of inequalities

$$x_1 \le \left(b_{n-1}^{(\upsilon)} + \sum_{i=1}^{n-1} b_{i-1}^{(\upsilon)} x_i \right) C(\lambda_0, \lambda, \gamma_1)$$

$$x_k \le b_{n-k}^{(\upsilon)} + \sum_{i=1}^{n-k} b_{i-1}^{(\upsilon)} x_{k+i-1} \quad , \quad k = 2, \ldots, n-1. \tag{7}$$

Denote by
$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \end{pmatrix},\ b = \begin{pmatrix} b_{n-2}^{(\upsilon)} \\ \vdots \\ b_0^{(\upsilon)} \end{pmatrix},$$

$$B = \begin{pmatrix} Cb_0^{(\upsilon)} & Cb_1^{(\upsilon)} & \cdots & Cb_{n-2}^{(\upsilon)} \\ 0 & b_0^{(\upsilon)} & \cdots & b_{n-3}^{(\upsilon)} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & b_0^{(\upsilon)} \end{pmatrix},\ C = C(\lambda_0, \lambda, \gamma_1).$$

We can rewrite this system of inequalities in the matrix form as

$$X \le b + BX. \tag{8}$$

Using the inequality $b_{n-i}^{(\upsilon)} \cdot b_i^{(\upsilon)} \le b_0^{(\upsilon)} \cdot b_n^{(\upsilon)} \cdot C_n^i$, $i = 0, 1, \ldots, n$, we obtain

$$Bb \le b_0^{(\upsilon)} \left(2^{n-1} - 1\right) \cdot Z \cdot b,$$

where

$$Z = \begin{pmatrix} C & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

and

$$X \le b + Bb + B^2 b + \ldots \le \sum_{n=0}^{\infty} \left[b_0^{(\upsilon)} \left(2^{n-1} - 1\right) \right]^n \cdot Z \cdot b.$$

Hence, we get

$$q = x_1 \le \frac{b_{n-1}^{(\upsilon)} \cdot C(\lambda_0, \lambda, \gamma_1)}{1 - b_0^{(\upsilon)} \left(2^{n-1} - 1\right)}. \tag{9}$$

The left-hand side inequality for q is obvious.

Lemma 2. If $b_0^{(\upsilon)}\left(2^{n-1}-1\right)\to 0$, then $q \sim C(\lambda_0,\lambda,\gamma_1)b_{n-1}^{(\upsilon)}$.

Proof. It follows immediately from the Lemma 1.

Let ς be time of a regeneration period of the random process $v(t)$. From the above Lemmas and the limiting theorem (3) we conclude the following theorem:

Theorem. Let $b_0^{(\upsilon)}\left(2^{n-1}-1\right)\to 0$ and $b_{n-1}^{(\upsilon)}\dfrac{E\varsigma^2}{(E\varsigma)^2}C(\lambda_0,\lambda,\gamma_1)\to 0$. Then the following asymptotic formula for the random variable τ_0 holds:

$$P\left(\lambda b_{n-1}^{(\upsilon)}\tau_0 > x\right)\to e^{-x}. \tag{10}$$

Acknowledgements

The authors wish to express their gratitude to A.D.Solovjev for formulation of the problem.

References

[1] A.Y.Khintchine: *About the mean time of machine interference.* Mat. Sb., 40 (1933), 119 – 123 (in Russian).

[2] C.Palm: *The Distribution of Repairmen in Servicing Automatic Machines.* Industritidn. Norden, 75 (1947), 75 – 80, 90 – 94, 119 – 123 (in Swedish).

[3] J.Taylor, R.R.P.Jackson: *An Application of the Birth and Death Process to the Provision of Spare Machines.* Operational Research Quart., 5 (1954), 95 – 108.

[4 P.Naor: *Normal Approximation to Machine Interference with Many Repairmen.* Journ. Roy. Stat. Soc., Ser. B, 19 (1957), 334 – 341.

[5] D.Gaver: *A Probability Problem Arising in Reliability and Traffic Studies.* Operations Research, 12 (4) (1964).

[6] C.Mack, K.Stoodley: *Machine Interference with Two Repairmen when Repair Time is Constant.* New Journ. Statist. and Oper. Res., 4 (2) (1968).

[7] M.Hollander, F.Proschan: *Testing whether new is better than used.* Ann. Math. Statist., 43 (4) (1972), 1136 – 1146.

[8] Mathematical problems of reliability theory. Radio i svyaz, Moscow, 1983 (in Russian).

[9] D.R.Cox: *Renewal Theory.* Methuen, Wiley, Lnd. – N.Y., 1962.

ÖZET

Bir tekrar üreme periodunda sistemin bozulması olasılığı q olsun. Sistemin bütün birimleri çalışır durumda iken sistemin bozuluncaya kadar geçen süre ise τ olsun. Güvenirliğin bu karakteristikleri incelenmiştir.

İstatistik, Journal of the Turkish Statistical Association
2000, Vol. 3, No 3, pp. 99-110

DISTRIBUTION OF RECORD TIMES FOR RANDOM WALK PROCESSES

I. Bairamov and Yilmaz Akdi
Ankara University, Department of Statistics
06100, Tandogan-Ankara

Abstract

Many time series as they occur in practice are not stationary. For example, the economies of many countries are developing or growing. The typical economic indicators will be showing a "trend" through time. This trend may be in the mean, the variance, or both. Such nonstationary time series are sometimes called evolutionary. In this study, the distribution of records for the first order nonstationary autoregressive processes are investigated. The upper records are considered but the same theory for the lower records can be carried out similarly by considering the sequence of random variables as $\{-X_i : i = 1, 2, 3, ...\}$. The distribution of the record times is derived and it is shown that record times have a Markovian property.

1. Introduction

Let $X_1, X_2, ..., X_n, ...$ be a sequence of random variables and

$$M(n) = Max(X_1, X_2, ..., X_n), \quad n = 1, 2, 3, ...$$

Define the record times as follows:

$$U(1) = 1, \quad U(n+1) = \min\left(j : j > U(n), X_j > X_{U(n)}\right), \; n = 1, 2, 3, ...$$

Let $X_{U(n)} = M(U(n))$ be n th record value, $n = 1, 2, 3, ...$By convention X_1 is a record value.

Developments of record theory have been reviewed by many authors including Galambos (1982), Nevzorov (1988), Nagaraja (1988), Arnold and Balakrishnan (1989), Arnold, Balakrishnan and Nagaraja (1992), Ahsanullah (1995).

It is well known that for the sequences of independent and identically distributed (i.i.d.) random variables $X_1, X_2, ..., X_n, ...$ with distribution function (d.f.) F the sequence of record times $\{U(n), n \geq 1\}$ form a Markov chain with transient probabilities

$$P\{U(n) = k | U(n-1) = j\} = \begin{cases} \frac{j}{k(k-1)} & k > j \geq n-1 \geq 2 \\ 0 & \text{for any other values of } j \text{ and } k \end{cases}$$

For independent and identically distributed random variables with distribution function F and probability density function f the distribution of nth record value is

$$F_n(x) = P\left\{X_{U(n)} \leq x\right\}$$

$$= \frac{1}{(n-1)!} \int_0^{\ln(1-F(x))} t^{n-1} e^{-t} dt = \frac{1}{(n-1)!} \int_0^x \left(\ln \frac{1}{1-F(x)}\right)^{n-1} f(x) dx$$

the probability density function of $X_{U(n)}$ is

$$f_n(x) = \frac{1}{(n-1)!} \left(\ln \frac{1}{1-F(x)}\right)^{n-1} f(x)$$

The joint probability density function of n records $X_{U(1)}, X_{U(2)}, ..., X_{U(n)}$ is given by

$$f(x_1, x_2, ..., x_n) = \begin{cases} \frac{f(x_1)}{1-F(x_1)} \cdots \frac{f(x_{n-1})}{1-F(x_{n-1})} f(x_n) & -\infty < x_1 < ... < x_n < \infty \\ 0 & \text{otherwise} \end{cases}$$

The joint probability density function of $X_{U(i)}$ and $X_{U(j)}$, $(i < j)$, is

$$f_{ij}(x_i, x_j) = \frac{(R(x_i))^{i-1}}{(i-1)!} \frac{f(x_i)}{1-F(x_i)} \frac{[R(x_j) - R(x_i)]^{j-i-1}}{(j-i-1)!} f(x_j), \ -\infty < x_1 < ... < x_n < \infty$$

where $R(x) = -\ln(1 - F(x))$.

Let $\xi_1 \equiv 1$ and $\xi_n = I_{\{M(n) > M(n-1)\}}$ for $n = 2, 3, 4, ...$ That is, ξ_n's are indicators of upper records. It is known that for independent and identically distributed random variables the record indicators ξ_n, $n > 1$ are independent random variables and $P\{\xi_n = 1\} = \frac{1}{n}$ for $n = 1, 2, 3, ...$ (see, Renyi (1962)). As noted by Nevzorova, Nevzorov and Balakrishnan (1997) this property of record indicators also holds for symmetrically dependent random variables.

Nevzorov (1985) introduced the so-called F^α scheme, which is defined as a sequence of independent random variables $\{X_i\}$ with distribution functions $F_i(x) = \{F(x)\}^{\alpha_i}$, for $i = 1, 2, 3, ...$ where $F(x)$ is a continuous distribution function and $\alpha_1, \alpha_2, ...$ are positive constants.

Nevzorov (1985) has shown that the record indicators $\xi_n, n \geq 1$ are also independent for the F^α scheme. Ballerini and Resnick (1987) consider a new record model in which the random variables $\{X_i\}$ can be dependent but the joint distribution of maxima $M_1, M_2, ...M_n$ coincides for any $n = 1, 2, 3, ...$ with the joint distribution of the corresponding vector for some F^α scheme. The record indicators $\xi_n, n \geq 1$ are also independent for Ballarini-Resnick scheme.

Ballerini (1994) considered the general class of dependent random variables, so-called Archimedian copula process with independent record indicators. A sequence of random variables $\{X_i\}$ with marginal distribution functions $\{F_i\}$ is said to be an Archimedian copula (AC) if for any $n = 1, 2, 3, ...,$ we obtain

$$P\{X_1 < t_1, ..., X_n < t_n\} = B\left(\sum_{i=1}^{n} A(F_i(t_i))\right),$$

where B is a monotone dependence function such that $B(0) = 1$ and $A = B^{-1}$ is the inverse of the dependence function B.

Note that the independence of record indicators $\xi_n, n \geq 1$ characterize the $\{F^\alpha\}$ scheme. There are some papers in this direction. For example see Borokov and Pfeifer (1995), Nevzorov (1993). But the mathematical theory of records has not been worked when the independence and identically distributiveness of the original random variables are discarded. Moreover, the stationarity condition of the original series is also removed.

Time series have many applications in different fields of sciences such as economics, engineering. In economics, the recorded history of economy is often in the form of the autoregressive models. Economic behavior is quantified in such a series as consumer price index, unemployment, gross national product, population, income, and consumption. In the area of time series, $X_n = \rho X_{n-1} + \epsilon_n$ has been considered and many testing procedures have been developed for testing $H_0 : \rho = 1$ against stationary alternatives. A practical and most popular one is the regression approach. Dickey and Fuller (1979) used a regression approach for testing the null hypothesis $H_0 : \rho = 1$ against stationary alternatives and they give the critical values of the test statistics. In Figure 1, IBM daily stock price, its autocorrelations and partial autocorrelations are plotted. From the time plot, it is seen that the series is increasing over time. This tells that the series is nonstationary in some way. Some tests are applied to this series and concluded that the process is nonstationary. We do not repeat their tests but instead directly look at the identification plots which will give us a rough idea of the order and stationarity of the series. We see that the decay of the autocorrelations are very slow and the partial autocorrelation function has a large spike (nearly 1) at the first lag, and others are small. We also look at the first difference of the original series. From SAS's PROC UNIVARIATE, we see that the first difference series look like a white noise sequence. That is, $X_n = X_{n-1} + \epsilon_n$ (ϵ_n are independent and identically distributed random variables with mean zero and constant variance σ^2) fit is appropriate for the IBM data.

Looking at the theoretical model, the random variables $X_1, X_2, ...X_n, ...$ are not independent and identically distributed, neither. Moreover, the process is not stationary (in the wide sense) because $Var(X_n) = n\sigma^2$ when $X_0 = 0$. In this paper we will study the distribution of the record times when the random variables form the model

$$X_n = X_{n-1} + \epsilon_n, \qquad \epsilon_n \ \ i.i.d. \ \ (0, \sigma^2)$$

2. Record Indicators

As noted earlier, for some certain class of models the record indicators $\{\xi_n, n \geq 1\}$ are independent. That is the events $\{X_i \text{ is a record}\}$ and $\{X_j \text{ is a record}\}$ are independent. However, in our case it will be shown that the record indicators are dependent random variables. Consider the first order autoregressive model

$$X_n = X_{n-1} + \epsilon_n, \ n = 1, 2, 3, ... \tag{1}$$

where ϵ_n's are independent and identically distributed random variables with mean zero and variance σ^2. Consider the random variable ν which is defined as follows: $\nu = n$ iff $\{S_1 \leq 0, S_2 \leq 0, ..., S_{n-1} \leq 0, S_n > 0\}$, where $S_n = \sum_{j=1}^{n} \epsilon_j$. Feller (1971) gives the form of the distribution of r.v. ν :

$$P\{\nu = n\} = P\{S_1 \leq 0, S_2 \leq 0, ..., S_{n-1} \leq 0, S_n > 0\} \equiv \tau_n, \quad n = 1, 2, 3, ...$$

The probability generating function of the r.v. ν is

$$\tau(s) = \sum_{n=1}^{\infty} \tau_n s^n \quad , \quad 0 \leq s \leq 1.$$

The distribution $\{\tau_n\}$ is completely determined by the probabilities $P\{S_n > 0\}$ and vise versa as a result of the following identity (see Feller 1971, p.413)

$$\log \frac{1}{1 - \tau(s)} = \sum_{n=1}^{\infty} \frac{s^n}{n} P\{S_n > 0\}$$

The assertion remains valid if the signs $>$ and $\leq$ are replaced by $\geq$ and $<$, respectively. These results are valid for any random variables $\epsilon_1, \epsilon_2, ...$ But when we assume that ϵ_n's are symmetrically distributed, then the result is simplified as

$$\tau_n = \frac{(2n)!}{(2n-1)(n!)^2 2^{2n}}, \quad n = 1, 2, 3, ... \tag{2}$$

Assume that $\epsilon_1, \epsilon_2, ..., \epsilon_n, ...$ is a sequence of independent and identically distributed random variables with a symmetric distribution function F, mean zero and variance σ^2.

Consider the model (1) $X_n = X_{n-1} + \epsilon_n$ where the random variable X_0 has a distribution function F_0. Note that $X_n = X_0 + \sum_{j=0}^{n} \epsilon_j = X_0 + S_n$ where $S_n = \epsilon_1 + \epsilon_2 + \ldots + \epsilon_n$. Define the random variable

$$M(n) = Max(X_1, X_2, \ldots, X_n), \quad n = 1, 2, 3, \ldots$$

and the record indicators as follows:

$$\xi_1 \equiv 1, \quad \xi_n = I_{\{M(n) > M(n-1)\}}, \quad n = 2, 3, 4, \ldots$$

That is, $\xi_n = 1$ if X_n is a record and $\xi_n = 0$ otherwise. By convention $P\{\xi_1 = 1\} = 1$ and some probabilities can be calculated as follows:

$$P\{\xi_2 = 1\} = P\{X_2 > X_1\} = P\{X_0 + \epsilon_1 + \epsilon_2 > X_0 + \epsilon_1\} = P\{\epsilon_2 > 0\} = \frac{1}{2}$$

and similarly

$$P\{\xi_3 = 1\} = P\{X_3 > X_2, X_3 > X_1\} = P\{X_0 + \epsilon_1 + \epsilon_2 + \epsilon_3 > X_0 + \epsilon_1 + \epsilon_2,$$

$$X_0 + \epsilon_1 + \epsilon_2 + \epsilon_3 > X_0 + \epsilon_1\} = P\{\epsilon_3 > 0, \epsilon_2 + \epsilon_3 > 0\}$$

$$= P\{\epsilon_3 > 0\} - P\{\epsilon_3 > 0, \epsilon_2 + \epsilon_3 < 0\} = \frac{1}{2} - \tau_2 = \frac{3}{8}$$

Analogously one can write

$$P\{\xi_4 = 1\} = P\{\epsilon_4 > 0, \epsilon_3 + \epsilon_4 > 0\} - P\{\epsilon_4 > 0, \epsilon_3 + \epsilon_4 > 0, \epsilon_2 + \epsilon_3 + \epsilon_4 < 0\}$$

$$= \frac{1}{2} - \tau_2 - \tau_3$$

and for a general case, it is easy to see that

$$P\{\xi_n = 1\} = \frac{1}{2} - \tau_2 - \tau_3 - \ldots - \tau_{n-1}$$

where

$$\tau_n = P\{\epsilon_1 < 0, \epsilon_1 + \epsilon_2 < 0, \ldots, \epsilon_1 + \epsilon_2 + \ldots + \epsilon_{n-1} < 0, \epsilon_1 + \epsilon_2 + \ldots + \epsilon_{n-1} > 0\}$$

$$= \frac{(2n)!}{(2n-1)(n!)^2 2^{2n}}$$

Moreover without having any difficulties one can calculate some joint probabilities as follows:

$$P\{\xi_1 = 1, \xi_2 = 1\} = P\{X_2 > X_1\} = \frac{1}{2},\ P\{\xi_1 = 1, \xi_2 = 0\} = P\{X_2 < X_1\} = \frac{1}{2},$$

$$P\{\xi_2 = 1, \xi_3 = 1\} = P\{X_2 > X_1, X_2 < X_3\} = P\{\epsilon_2 > 0, \epsilon_3 > 0\} = \frac{1}{4}$$

Since

$$P\{\xi_2 = 1, \xi_3 = 1\} = \frac{1}{4} \neq P\{\xi_2 = 1\}P\{\xi_3 = 1\} = \frac{1}{2}\frac{3}{8} = \frac{3}{16}$$

the record indicators ξ_n's are not independent. That is, the events $\{X_i \text{ is a record}\}$ and $\{X_j \text{ is a record}\}$ are dependent.

3. Distribution of $U(2)$

In this section, we will derive the distribution of the second record time. This is going to simplify finding the distribution of the n th record time. Now we state the following theorem.

Theorem 1. Consider the nonstationary first order autoregressive time series model $X_n = X_{n-1} + \epsilon_n$, $n = 1, 2, 3, ...$ where ϵ_n's are independent and identically distributed random variables with mean zero and variance σ^2. If ϵ_n's have a symmetric and continuous distribution function F, then the distribution of the second record time is

$$P\{U(2) = n\} = \tau_{n-1}\ ,\ n = 2, 3, 4, ... \tag{3}$$

where τ_n is defined in (2).

Proof. The event that the second record time being n is the event that the value of the first random variable is larger that all the values of the random variables

$X_2, X_3, ..., X_{n-1}$ but it is smaller than the value of the random variable X_n. That is, In order to find the distribution of the second record time we need to calculate the following:

$$P\{U(2) = n\} = P\{X_1 > X_2, X_1 > X_3, ..., X_1 > X_{n-1}, X_1 < X_n\}$$

$$= P\{X_0 + \epsilon_1 > X_0 + \epsilon_1 + \epsilon_2, X_0 + \epsilon_1 > X_0 + \epsilon_1 + \epsilon_2 + \epsilon_3, ...,$$

$$X_0 + \epsilon_1 > X_0 + \epsilon_1 + ... + \epsilon_{n-1}, X_0 + \epsilon_1 < X_0 + \epsilon_1 + ... + \epsilon_n\}$$

$$= P\{\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{n-1} < 0, \epsilon_2 + ... + \epsilon_n > 0\}$$

$$= P\{\epsilon_1 < 0, \epsilon_1 + \epsilon_2 < 0, ..., \epsilon_1 + ... + \epsilon_{n-2} < 0, \epsilon_2 + ... + \epsilon_{n-1} > 0\} = \tau_{n-1}$$

which completes the proof of the theorem.◇

Theorem 2. The sequence of random variables $U(1), U(2), U(3), ..., U(n), ...$ form a Markov chain with transient probabilities τ_{j-i} for $j > i$. That is, the conditional probability of $U(n)$ being at the state j does not depend on the past states except the state of $U(n-1)$. In other words, when $j > i$ we have the following:

$$P\{U(n) = j | U(n-1) = i, U(n-2) = i_{n-2}, ..., U(3) = i_3, U(2) = i_2\}$$

$$= P\{U(n) = j | U(n-1) = i\} = \begin{cases} \tau_{j-i} & \text{if } j > i. \\ 0 & \text{if } j \leq i \end{cases} \qquad (4)$$

Proof. Using the conditional probability formula

$$P\{U(n) = j | U(n-1) = i, U(n-2) = i_{n-2}, ..., U(3) = i_3, U(2) = i_2\}$$

$$= \frac{P\{U(n) = j, U(n-1) = i, U(n-2) = i_{n-2}, ..., U(3) = i_3, U(2) = i_2\}}{P\{U(n-1) = i, U(n-2) = i_{n-2}, ..., U(3) = i_3, U(2) = i_2\}} = \frac{\alpha}{\beta}$$

where

$$\alpha = P\{U(n) = j, U(n-1) = i, U(n-2) = i_{n-2}, ..., U(3) = i_3, U(2) = i_2\}$$

$$= P\{X_1 > X_2, X_1 > X_3, ..., X_1 > X_{i_2-1}, X_1 < X_{i_2}; X_{i_2} > X_{i_2+1}, ..., \\ X_{i_2} > X_{i_3-1}, X_{i_2} < X_{i_3}; ..., X_{i_{n-2}} > X_{i_{n-2}+1}, ... X_{i_{n-2}} > X_{i-1}, \\ X_{i_{n-2}} < X_i; X_i > X_{i+1}, X_i > X_{i+2}, ..., X_i > X_{j-1}, X_i < X_j\}$$

$$= P\{X_0 + S_1 > X_0 + S_2, ..., X_0 + S_1 > X_0 + S_{i_2-1}, X_0 + S_1 < X_0 + S_{i_2};$$
$$X_0 + S_{i_2} > X_0 + S_{i_2+1}, ..., X_0 + S_{i_2} > X_0 + S_{i_3-1}, X_0 + S_{i_2} < X_0 + S_{i_3}; ...;$$
$$X_0 + S_{i_{n-2}} > X_0 + S_{i_{n-2}+1}, ..., X_0 + S_{i_{n-2}} > X_0 + S_{i-1}, X_0 + S_{i_{n-2}} < X_0 + S_i;$$
$$X_0 + S_i > X_0 + S_{i+1}, ..., X_0 + S_i > X_0 + S_{j-1}, X_0 + S_i < X_0 + S_j\}$$

$$= P\{\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{i_2-1} < 0, \epsilon_2 + ... + \epsilon_{i_2} > 0;$$

$$\epsilon_{i_2+1} < 0, \epsilon_{i_2+1} + \epsilon_{i_2+2} < 0, ..., \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} < 0, \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} > 0; ...;$$

$$\epsilon_{i_{n-2}+1} < 0, \epsilon_{i_{n-2}+1} + \epsilon_{i_{n-2}+2} < 0, ..., \epsilon_{i_{n-2}+1} + ... + \epsilon_{i-1} < 0, \epsilon_{i_{n-2}+1} + ... + \epsilon_i > 0;$$

$$\epsilon_{i+1} < 0, \epsilon_{i+1} + \epsilon_{i+2} < 0, ..., \epsilon_{i+1} + ... + \epsilon_{j-1} < 0, \epsilon_{i+1} + ... + \epsilon_{j-1} > 0\}$$

Since ϵ_n's are independently distributed random variables, this probability can be written as

$$= P\{\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{i_2-1} < 0, \epsilon_2 + ... + \epsilon_{i_2} > 0\}$$
$$P\{\epsilon_{i_2+1} < 0, \epsilon_{i_2+1} + \epsilon_{i_2+2} < 0, ..., \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} < 0, \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} > 0\}$$
$$.....$$
$$P\{\epsilon_{i_{n-2}+1} < 0, \epsilon_{i_{n-2}+1} + \epsilon_{i_{n-2}+2} < 0, ..., \epsilon_{i_{n-2}+1} + ... + \epsilon_{i-1} < 0, \epsilon_{i_{n-2}+1} + ... + \epsilon_i > 0\}$$
$$P\{\epsilon_{i+1} < 0, \epsilon_{i+1} + \epsilon_{i+2} < 0, ..., \epsilon_{i+1} + ... + \epsilon_{j-1} < 0, \epsilon_{i+1} + ... + \epsilon_j > 0\}$$

Using the same argument, the probability in the denominator can be calculated as

$$\beta = P\{\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{i_2-1} < 0, \epsilon_2 + ... + \epsilon_{i_2} > 0\}$$
$$P\{\epsilon_{i_2+1} < 0, \epsilon_{i_2+1} + \epsilon_{i_2+2} < 0, ..., \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} < 0, \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} > 0\}$$
$$...$$
$$P\{\epsilon_{i_{n-2}+1} < 0, \epsilon_{i_{n-2}+1} + \epsilon_{i_{n-2}+2} < 0, ..., \epsilon_{i_{n-2}+1} + ... + \epsilon_{i-1} < 0, \epsilon_{i_{n-2}+1} + ... + \epsilon_i > 0\}$$

and thus the conditional probability in the left hand side of (4) is

$$P\{U(n) = j | U(n-1) = i, U(n-2) = i_{n-2}, ..., U(3) = i_3, U(2) = i_2\}$$
$$= P\{\epsilon_{i+1} < 0, \epsilon_{i+1} + \epsilon_{i+2} < 0, ..., \epsilon_{i+1} + ... + \epsilon_{j-1} < 0, \epsilon_{i+1} + ... + \epsilon_j > 0\} \qquad (5)$$
$$= \tau_{j-i}$$

In order to calculate the right hand side of (4), we use the same argument. Without repeating the same calculations, the probability on the right hand side of (4) is

$$P\{U(n) = j | U(n-1) = i\}$$
$$= P\{\epsilon_{i+1} < 0, \epsilon_{i+1} + \epsilon_{i+2} < 0, ..., \epsilon_{i+1} + ... + \epsilon_{j-1} < 0, \epsilon_{i+1} + ... + \epsilon_j > 0\} \qquad (6)$$
$$= \tau_{j-i}$$

The identities in (5) and (6) imply that the sequence of random variables $U(1)$, $U(2), ..., U(n), ...$ form a Markov chain with transient probabilities τ_{j-i} and this completes the proof.◇

To clarify the results, we take $n = 4$ and show that

$$P\{U(4) = i_4 | U(3) = i_3, U(2) = i_2\} = P\{U(4) = i_4 | U(3) = i_3\} \tag{7}$$

The right hand side of (7) is

$$P\{U(4) = i_4 | U(3) = i_3, U(2) = i_2\}. = \frac{\alpha_1}{\beta_1}$$

where

$$\alpha_1 = P\{U(4) = i_4, U(3) = i_3, U(2) = i_2\}$$

$$\begin{aligned} = P\{&X_1 > X_2, X_1 > X_3, ..., X_1 > X_{i_2-1}, X_1 < X_{i_2};\\ &X_{i_2} > X_{i_2+1}, X_{i_2} > X_{i_2+2}, ..., X_{i_2} > X_{i_3-1}, X_{i_2} < X_{i_3};\\ &X_{i_3} > X_{i_3+1}, X_{i_3} > X_{i_3+2}, ..., X_{i_3} > X_{i_4-1}, X_{i_3} < X_{i_4}\} \end{aligned}$$

$$\begin{aligned} = P\{&\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{i_2-1} < 0, \epsilon_2 + ... + \epsilon_{i_2} > 0;\\ &\epsilon_{i_2+1} < 0, \epsilon_{i_2+1} + \epsilon_{i_2+2} < 0, ..., \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} < 0, \epsilon_{i_2+1} + ... + \epsilon_{i_3} > 0;\\ &\epsilon_{i_3+1} < 0, \epsilon_{i_3+1} + \epsilon_{i_3+2} < 0, ..., \epsilon_{i_3+1} + ... + \epsilon_{i_4-1} < 0, \epsilon_{i_3+1} + ... + \epsilon_{i_4} > 0\} \end{aligned}$$

$$\begin{aligned} &= P\{\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{i_2-1} < 0, \epsilon_2 + ... + \epsilon_{i_2} > 0\}\\ &P\{\epsilon_{i_2+1} < 0, \epsilon_{i_2+1} + \epsilon_{i_2+2} < 0, ..., \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} < 0, \epsilon_{i_2+1} + ... + \epsilon_{i_3} > 0\}\\ &P\{\epsilon_{i_3+1} < 0, \epsilon_{i_3+1} + \epsilon_{i_3+2} < 0, ..., \epsilon_{i_3+1} + ... + \epsilon_{i_4-1} < 0, \epsilon_{i_3+1} + ... + \epsilon_{i_4} > 0\} \end{aligned}$$

$$= \tau_{i_2-1}\tau_{i_3-i_2}\tau_{i_4-i_3}$$

and

$$\beta_1 = P\{U(3) = i_3, U(2) = i_2\}$$

$$\begin{aligned} = P\{&X_1 > X_2, X_1 > X_3, ..., X_1 > X_{i_2-1}, X_1 < X_{i_2};\\ &X_{i_2} > X_{i_2+1}, X_{i_2} > X_{i_2+2}, ..., X_{i_2} > X_{i_3-1}, X_{i_2} < X_{i_3}\} \end{aligned}$$

$$\begin{aligned} &= P\{\epsilon_2 < 0, \epsilon_2 + \epsilon_3 < 0, ..., \epsilon_2 + ... + \epsilon_{i_2-1} < 0, \epsilon_2 + ... + \epsilon_{i_2} > 0\}\\ &P\{\epsilon_{i_2+1} < 0, \epsilon_{i_2+1} + \epsilon_{i_2+2} < 0, ..., \epsilon_{i_2+1} + ... + \epsilon_{i_3-1} < 0, \epsilon_{i_2+1} + ... + \epsilon_{i_3} > 0\} \end{aligned}$$

$$= \tau_{i_2-1}\tau_{i_3-i_2}$$

That is the conditional probability is

$$P\{U(4) = i_4 | U(3) = i_3, U(2) = i_2\}. = \frac{\alpha_1}{\beta_1}$$

$$= \frac{\tau_{i_2-1}\tau_{i_3-i_2}\tau_{i_4-i_3}}{\tau_{i_2-1}\tau_{i_3-i_2}} = \tau_{i_4-i_3}$$

In order to calculate the probability on the left hand side of (7) we will use the total probability formula. The left hand side of (7) is

$$P\{U(4) = i_4 | U(3) = i_3\} = \frac{\alpha_2}{\beta_2}$$

where

$$\alpha_2 = P\{U(4) = i_4, U(3) = i_3\} = \sum_{i_2=2}^{\infty} P\{U(4) = i_4, U(3) = i_3, U(2) = i_2\}$$

the probability in the sums is calculated in obtaining the probability $\alpha_1 = \tau_{i_2-1}\tau_{i_3-i_2}\tau_{i_4-i_3}$ and thus

$$\alpha_2 = P\{U(4) = i_4, U(3) = i_3\} = \sum_{i_2=2}^{\infty} \tau_{i_2-1}\tau_{i_3-i_2}\tau_{i_4-i_3} = \tau_{i_4-i_3} \sum_{i_2=2}^{\infty} \tau_{i_2-1}\tau_{i_3-i_2}$$

and similarly

$$\beta_2 = P\{U(3) = i_3\} = \sum_{i_2=2}^{\infty} P\{U(3) = i_3, U(2) = i_2\} = \sum_{i_2=2}^{\infty} \tau_{i_2-1}\tau_{i_3-i_2}$$

and thus we have

$$RHS = \frac{\alpha_1}{\beta_1} = \tau_{i_4-i_3} = LHS = \frac{\alpha_2}{\beta_2} = \frac{\tau_{i_4-i_3} \sum_{i_2=2}^{\infty} \tau_{i_2-1}\tau_{i_3-i_2}}{\sum_{i_2=2}^{\infty} \tau_{i_2-1}\tau_{i_3-i_2}}$$

and hence the equality holds.

The main result of this study follows from the Theorems 1 and Theorem 2

Theorem 3. The distribution of the n th record time is

$$F_n(j) = P\{U(n) = j\} = \sum_{k=n-1}^{j-1} F_n(j)\tau_{j-i}.$$

Proof. Using the Theorem 2 and Theorem 3 with the formula of total probability concludes the proof:

$$\begin{aligned} F_n(j) = P\{U(n) = j\} &= \sum_{k=n-1}^{j-1} P\{U(n) = j | U(n-1) = k\} P\{U(n-1) = k\} \\ &= \sum_{k=n-1}^{j-1} F_n(j)\tau_{j-i}. \end{aligned} \qquad (8)$$

To calculate the distribution of the third record time we write

$$P\{U(3) = k\} = \sum_{i=2}^{k-1} P\{U(3) = k|U(2) = i\}P\{U(2) = i\} = \sum_{i=2}^{k-1} \tau_{k-i}\tau_{i-1}$$

rest of the probabilities can be calculated recursively.

References

[1] Ahsanullah M.(1995) *Record Statistics.* Nova Science publishers, Inc., Commack,NY.

[2] Arnold B.C. , Balakrishnan, N. (1989) Relations, Bounds and Approximations for Order Statistics. *Lecture Notes in Statistics* No. **53,** New York: Springer-Verlag.

[3] Arnold B.C. , Balakrishnan, N., Nagaraja H.N. (1992) *A First Course of Order Statistics.* Wiley, New York.

[4] Ballerini R. (1994) A Dependent F^{α} scheme. *Statist. Probab. Letters.* **21**, 2,21-25.

[5] Ballerini R., Resnick S.I. (1987) Embedding Sequences of Successive Maxima in Extremal Processes, with Applications. *J. Appl. Probab.*, **24**,827-837.

[6] Borovkov K., Pfiefer D. (1995) On record Indices and Record Times. *J. Statist. Planning and Inference.* **45**,65-80.

[7] Dickey, D.A. , Fuller, W.A. (1979), *Distribution of the Estimators for Autoregressive Time series with a Unit Root,* Journal of the American Statistical Association, **74**, 427-431

[8] Galambos J.(1978) *The Asymptotic Theory of Extreme Order Statistics.* Wiley.New York.

[9] Nagaraja H.N.(1988) Record Values and Related Statistics. A Review. *Commun in Statist. Theor Meth.* **17**, 2223-2238.

[10] Nevzorov V.B.(1988) Records. *Theor.Prob.Apl.* **32**,201-228.

[11] Nevzorov V.B. (1987) Record and Inter Record Times for Sequences of Non-identically Distributed Random Variables. *Journal of Soviet Mathematics*,**36**,510-516.

[12] Nevzorov V.B. (1993) *Characterizations of Certain Nonstationary Sequences by Properties of Maxima and Records.* In Rings and Modules: Limit Theorems in Probability Theory (Z.I.Borevich and V.V.Petrov Eds),Vol.**3**, 188-197.

[13] Renyi A. (1976) On The Extreme Elements of Observations. *Selected Papers of Alfred Renyi.*3,50-65. Budapest, Akademiai Kiado.

ÖZET

Bu çalışmada durağan olmayan birinci dereceden otoregressiv zaman serileri için rekor değerlerin dağılımları incelenmiştir. Bu bağımlı rasgele degişkenler dizisi için rekor zamanlarının dağılımları elde edilmiş ve rekor zamanları dizisinin Markov zinciri oluşturduğu gösterilmiştir.

Figure 1: IBM Daily Stock Prices