REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000
https://doi.org/00.00000/revstat.v00i0.000

On the New Properties of Conditional Expectations

Authors: Ismihan Bayramoglu ✉
– Department of Mathematics, Izmir University of Economics, Izmir, Turkey
ismihan.bayramoglu@ieu.edu.tr

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- The concept of conditional expectation is important in applications of probability and statistics in many areas such as reliability engineering, economy, finance, and actuarial sciences due to its property of being the best predictor of a random variable as a function of another random variable. This concept also is essential in the martingale theory and theory of Markov processes. Even though, there has been studied and published many interesting properties of conditional expectations with respect to a sigma-algebra generated by a random variable it remains an attractive subject having interesting applications in many fields. In this paper, we present some new properties of the conditional expectation of a random variable given another random variable. The copula and dependence properties of conditional expectations as random variables are also studied. We present also some new inequalities having interesting applications and results in martingale theory and Markov processes.

Keywords:

- *Conditional expectation; sigma algebra; order statistics; prediction.*

AMS Subject Classification:

- 62H20, 62G30.

✉ Corresponding author

1. INTRODUCTION

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Consider the random variables X and Y defined in this probability space and having joint distribution function $F_{X,Y}(x,y) = C(F_X(x), F_Y(y)), (x,y) \in \mathbb{R}^2$, where $C(u,v), (u,v) \in I^2 \equiv [0,1]^2$ is a connected copula (see Nelsen ([11])). Consider functions $\varphi(y) = E(X \mid Y = y), y \in \mathbb{R}$ and $\psi(x) = E(Y \mid X = x), x \in \mathbb{R}$ and random variables $Z_1 \equiv \varphi(Y) = E(X \mid Y)$ and $Z_2 \equiv \psi(X) = E(Y \mid X)$. It is well known that the best predictor of X by Y in the sense of least square distance is $\varphi(Y)$, and the best predictor of Y by X is $\psi(X)$, i.e.

$$\min_g E(X - g(Y))^2 = E(X - \varphi(Y))^2$$
$$\min_h E(Y - h(X))^2 = E(Y - \psi(X))^2,$$

where the min is taken over all measurable functions g and h. The conditional expectation $E(X \mid Y)$ is a random variable defined in the same probability space $(\Omega, \mathcal{F}, P)$. The random variable defined as a conditional expectation $E(X \mid Y)$ is an important classical concept, it is the best predictor for X as a function of Y, and plays a crucial role in many theoretical and practical aspects of probability theory. For example, in practical applications, if we know the joint distribuıtion of X and Y and the value of Y, we can use $\varphi(Y)$ instead of random variable X, whose values are very difficult, expensive, or impossible to measure. A wide description of the concept of conditional expectation and its properties can be found in many books on probability and statistics including Borovkov (1998) ([8]), Ross (2002) ([13]), among others. In this paper, we aim to consider some unknown and interesting properties of conditional expectations having applications in many areas such as economics, engineering, actuarial sciences, and financial mathematics.

The paper is organized as follows. In Section 2 we consider the conditional expectation of the random variable Y given X and compare it with the random variable defined as the arithmetic mean of conditional expectations of Y given $X_1, X_2, ..., X_n$ which are the copies (dependent or independent) of X. The application of the results in finance is shown. In Section 3 we are interested in the joint distribution of random variables $\varphi(Y)$ and $\psi(X)$ and study the dependence properties and copulas of these random variables. In Section 4 we consider the sequence of any random variables $X_1, X_2, ..., X_n, ...$ and study the properties of the sequence of random variables defined as $Y_1 = X_1, Y_2 = E(X_2 \mid X_1), ..., Y_n = E(X_n \mid X_{n-1}), ...$ and the sequence of random variables defined as $E(X_{n+1} \mid X_{i_1}, X_{i_2}, ..., X_{i_k}), 1 \le i_1 < i_2 < ... < i_k \le n, 1 \le k \le n.$ We present some theorems describing the interesting properties of these sequences and provide examples comparing them with Markov sequences and martingales.

2. SOME INEQUALITIES INVOLVING CONDITIONAL EXPECTATIONS

Let $X_1, X_2, ..., X_n$ be the copies of the random variable X. Let

$$\hat{Y} = E(Y \mid X)$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} E(Y \mid X_i).$$

The following theorem presents a simple inequality and makes it possible to compare the predicted value of Y through $\hat{Y}$ with the predicted value of Y through $\bar{Y}$.

Theorem 2.1. *Let X any Y be any random variables defined on the same propbability space and $X_1, X_2, ..., X_n$ be the copies of X, i.e random variables (dependent or independent) having the same distribution as X. Then*

$$E\left(Y - \frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \leq E(Y - X)^2.$$

Proof: Using Schwarz inequality we can write

$$\begin{aligned}
&E\left(Y - \frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \\
&= \frac{1}{n^2}E\left(\sum_{i=1}^{n}(Y - X_i)\right)^2 \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} E(Y - X_i)^2 + 2\sum_{1\leq i<j\leq n}^{n} E(Y - X_i)(Y - X_j)\right) \\
&\leq \frac{1}{n^2}\left(\sum_{i=1}^{n} E(Y - X_i)^2 + 2\sum_{1\leq i<j\leq n}^{n} \left(E(Y - X_i)^2\right)^{\frac{1}{2}}\left(E(Y - X_j)^2\right)^{\frac{1}{2}}\right) \\
&= \frac{1}{n^2}\left(nE(Y - X)^2 + \frac{n(n-1)}{2}2E(Y - X)^2\right) = E(Y - X)^2.
\end{aligned}$$

□

If $E(Y \mid X)$ is used instead of X in Theorem 2.1, then the following Corollary is obtained.

Corollary 2.1. *It is true that*

$$E(Y-\bar{Y})^2 \leq E(Y-\hat{Y})^2,$$

$$E\left(Y-\frac{1}{n}\sum_{i=1}^{n}E(Y\mid X_i)\right)^2 \leq E\left(Y-E(Y\mid X)\right)^2.$$

Remark 2.1. Considering

$$Eg\left(Y-\sum_{i=1}^{n}a_iX_i\right)$$

insted of

$$E\left(Y-\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2,$$

where g is any convex function and $\sum_{i=1}^{n}a_i=1$, and using Jensen's inequality for convex functions Theorem 2.1 can be extended as

$$g\left(E(Y-X)\right) \leq Eg\left(Y-\sum_{i=1}^{n}a_iX_i\right) \leq Eg(Y-X).$$

3. COPULA AND COVARIANCE

Since $E(X\mid Y)$ is the best predictor for X as a function of Y, and $E(Y\mid X)$ is the best predictor of Y as a function of X, it would be interesting to investigate how the dependence structure will change if we replaced X with $E(X\mid Y)$ and Y with $E(Y\mid X)$. For this purpose consider the joint distribution of the random variables $Z_1\equiv\varphi(Y)=E(X\mid Y)$ and $Z_2\equiv\psi(X)=E(Y\mid X)$. We are interested in copula of Z_1 and Z_2. Let $\varphi^{-1}(y)=\inf\{x:\varphi(x)\leq y\}$ and $\psi^{-1}(x)=\inf\{y:\psi(y)\leq x\}$ are the generalized inverses of φ and ψ. Consider the joint distribution function of Z_1 and Z_2. Let $F_X(x)$ and $F_Y(x)$ be a distribution function of X and Y, respectively. Assuming X and Y have the same support, denote left and right endpoints of the support of X and Y by $a=\inf\{x:F_X(x)>0\}$ and $b=\sup\{x:F_X(x)<1\}$, respectively. We allow also the cases $a=-\infty$ and $b=\infty$. We have

$$\begin{aligned}
F_{Z_1,Z_2}(z_1,z_2) &= P\{Z_1\leq z_1, Z_2\leq z_2\}\\
&= P\{\varphi(Y)\leq z_1,\psi(X)\leq z_2\}\\
&= P\{Y\leq\varphi^{-1}(z_1),X\leq\psi^{-1}(z_2)\} = P\{X\leq\psi^{-1}(z_2),Y\leq\varphi^{-1}(z_1)\}\\
&= F_{X,Y}(\psi^{-1}(z_2),\varphi^{-1}(z_1)),(z_1,z_2)\in[a,b]^2
\end{aligned}$$

Hereafter, we assume that $\psi^{-1}(\infty)=\infty$ and $\varphi^{-1}(\infty)=\infty$ and $\psi^{-1}(-\infty)=-\infty$ and $\varphi^{-1}(-\infty)=-\infty$. The marginal distributions are

$$F_{Z_1}(z_1)\equiv P\{Z_1\le z_1\}=\lim_{z_2\to\infty}P\{X\le\psi^{-1}(z_2),Y\le\varphi^{-1}(z_1)\}=F_Y(\varphi^{-1}(z_1))$$
$$F_{Z_2}(z_2)\equiv P\{Z_2\le z_2\}=\lim_{z_1\to\infty}P\{X\le\psi^{-1}(z_2),Y\le\varphi^{-1}(z_1)\}=F_X(\psi^{-1}(z_2)).$$

Now we are interested in the copula of random vector (Z_1,Z_2). Denote

$$F_X^{-1}(x)=\inf\{y:F_X(y)\ge x\}\text{ and }F_Y^{-1}(y)=\inf\{x:F_Y(x)\ge y\}.$$

Now, consider

$$F_{Z_1,Z_2}(z_1,z_2)=C_{Z_1,Z_2}(F_{Z_1}(z_1),F_{Z_2}(z_2)),\tag{3.1}$$

where $C_{Z_1,Z_2}(t,s)$ is a connecting copula of Z_1 and Z_2.

Using probability integral transformation

$$\begin{aligned}&F_{Z_1}(z_1)=t\ \Leftrightarrow F_Y(\varphi^{-1}(z_1))=t\ \Leftrightarrow z_1=\varphi(F_Y^{-1}(t)),\ F_{Z_1}^{-1}(t)=\varphi(F_Y^{-1}(t))\\ &F_{Z_2}(z_2)=s\ \Leftrightarrow F_X(\psi^{-1}(z_2))=s\ \Leftrightarrow z_2=\psi(F_X^{-1}(s)),F_{Z_2}^{-1}(s)=\psi(F_X^{-1}(s))\end{aligned}\tag{3.2}$$

we obtain from (3.1) and (3.2)

$$\begin{aligned}C_{Z_1,Z_2}(t,s)&=F_{Z_1,Z_2}(F_{Z_1}^{-1}(t),F_{Z_2}^{-1}(s))=F_{X,Y}(\psi^{-1}(z_2),\varphi^{-1}(z_1))\\&=F_{X,Y}(\psi^{-1}(\psi(F_X^{-1}(s)),\varphi^{-1}(\varphi(F_Y^{-1}(t))\\&=F_{X,Y}(F_X^{-1}(s),F_Y^{-1}(t))=C(s,t)\end{aligned}$$

Therefore, we can formulate the following theorem.

Theorem 3.1. *Let the joint distribution function of random variables X and Y be $F_{X,Y}(x,y)=C(F_X(x),F_Y(y)),(x,y)\in[a,b]^2$, where $C(u,v),(u,v)\in I^2\equiv[0,1]^2$ is a connected copula. Consider functions $\varphi(y)=E(X\mid Y=y),y\in\mathbb{R}$ and $\psi(x)=E(Y\mid X=x),x\in\mathbb{R}$ and random variables $Z_1\equiv\varphi(Y)=E(X\mid Y)$ and $Z_2\equiv\psi(X)=E(Y\mid X)$. Assume that $\lim_{t\to\infty}\psi^{-1}(t)=\infty$ and $\lim_{s\to\infty}\varphi^{-1}(s)=\infty$. Then the copula of Z_1 and Z_2 is $C_{Z_1,Z_2}(t,s)=C(s,t)$, $0\le t,s\le 1$. Therefore, if X and Y are exchangeable then $C_{Z_1,Z_2}(t,s)=C(t,s)$.*

Example 3.1. Let $(X;Y)$ be a bivariate normal random vector with joint pdf

$$\begin{aligned}f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}((\frac{x-\mu_1}{\sigma_1})^2\\-2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{x-\mu_2}{\sigma2})+(\frac{y-\mu_2}{\sigma_2})^2\}.\end{aligned}$$

Then

$$\psi(x) = E(Y \mid X = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) \tag{3.3}$$

$$\varphi(y) = E(X \mid Y = y) = \mu_1 + \rho\frac{\sigma_1}{\sigma 2}(y - \mu_2) \tag{3.4}$$

and

$$\psi^{-1}(t) = \frac{t - \mu_2}{\rho\sigma_2}\sigma_1 + \mu_1 \tag{3.5}$$

$$\varphi^{-1}(s) = \frac{s - \mu_1}{\rho\sigma_1}\sigma_2 + \mu_2 \tag{3.6}$$

and

$$\lim_{t\to\infty} \psi^{-1}(t) = \infty$$
$$\lim_{s\to\infty} \varphi^{-1}(s) = \infty.$$

Therefore, $C_1(x, y) = C(y, x)$, where C is a copula of (X,Y) and C_1 is a copula of $(\psi(Y), \varphi(X)) = (E(X \mid Y), E(Y \mid X))$.

We will use the following well-known property of the conditional expectation:

$$E(XY) = E(E(XY \mid Y)) = E(YE(X \mid Y))$$

Proposition 3.1.

$$Cov(E(X \mid Y), Y) = Cov(E(Y \mid X), X) = Cov(X, Y).$$

Proof: Since

$$E(XY) = E[E(XY \mid Y)] = E[YE(X \mid Y)]$$

and

$$E(Y)E[E(X \mid Y)] = E(Y)E(X)$$

then,

$$\begin{aligned} Cov(X, Y) &= E(XY) - E(X)E(Y) = E(YE(X \mid Y)) - E(Y)E(E(X \mid Y)) \\ &= Cov(Y, E(X \mid Y)) = Cov(E(X \mid Y), Y). \end{aligned}$$

□

Remark 3.1. It can be observed that $Cov(E(X \mid Y), E(Y \mid X))$ may not be equal to $Cov(X, Y)$.

Indeed,

$$\begin{aligned}
&E\psi(Y)\varphi(X) - E\psi(Y)E\varphi(X) \\
&= E\psi(Y)\varphi(X) - E[E(X \mid Y)]E[E(Y \mid X)] \\
&= E\psi(Y)\varphi(X) - E(X)E(Y) \\
&= E[E(X \mid Y)(E(X \mid Y)] - E(X)E(Y).
\end{aligned}$$

Let for example $\psi(Y) = aY + b$, $\varphi(X) = cX + d$, where $a, b, c, d > 0$ (see for example (3.3) and (3.4)). Then $E\psi(Y)\varphi(X) = acE(XY) + adEY + bcEX + bd$. Therefore, $E\psi(Y)\varphi(X) = acE(XY) + adEY + bcEX + bd = EXY$ only if $a = 1, c = 1, b = 0, d = 0$. For (3.3) and (3.4) this means that it must be $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 1$.

4. SEQUENCES OF PREDICTED RANDOM VARIABLES

Let $X_1, X_2, ..., X_n, ...$ be a sequence of dependent random variables. Let $Y_1 = X_1, Y_2 = E(X_2 \mid X_1), ..., Y_n = E(X_n \mid X_{n-1}), ...$ It is clear that $EY_i = E(E(X_i \mid X_{i-1})) = EX_i, i = 12, ...$ Since $E(Y_1 Y_2) = E(X_1 E(X_2 \mid X_1)) = E(E(X_1 X_2 \mid X_1)) = EX_1 X_2$, then $Cov(Y_1, Y_2) = Cov(X_1, X_2)$. Furthermore, denoting by $\psi_i(x) = E(X_i \mid X_{i-1} = x)$, $i = 2, 3, ...$, we have $\psi_i(X_{i-1}) = E(X_i \mid X_{i-1})$.

It is well known that the best predictor for X_{n+1} expressed as a function of $X_1, X_2, ..., X_n$ is $E(X_{n+1} \mid X_1, X_2, ..., X_n) = \Psi(X_1, X_2, ..., X_n)$, i.e.

$$\min_g E(X_{n+1} - g(X_1, X_2, ..., X_n))^2 = E(X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_n))^2$$

Theorem 4.1. *Let X, Y and Z be any random variables defined on probability space $\{\Omega, F, P\}$. Then*

$$E[X - E(X \mid Y, Z)]^2 \leq \min\left\{E[X - E(X \mid Y)]^2, E[X - E(X \mid Z)]^2\right\} \tag{4.1}$$

Proof: Consider

$$\begin{aligned}
&E\left[(X - E(X \mid Y, Z))^2 \mid Y, Z\right] \\
&= E[(X - E(X \mid Y) + E(X \mid Y) - E(X \mid Y, Z))^2 \mid Y, Z]
\end{aligned}$$

$$
\begin{aligned}
&= E[(X - E(X \mid Y))^2 \mid Y, Z] \\
&+ 2E\left[\, (X - E(X \mid Y))\, (E(X \mid Y) - E(X \mid Y, Z)) \mid Y, Z\right] \\
&+ E\left[(E(X \mid Y) - E(X \mid Y, Z))^2 \mid Y, Z\right] \\
&= E[\{X - E(X \mid Y)\}^2 \mid Y, Z] \\
&+ 2\,[E(X \mid Y) - E(X \mid Y, Z)]\, E\left[\{X - E(X \mid Y)\} \mid Y, Z\right] \\
&+ E\left[\{E(X \mid Y) - E(X \mid Y, Z)\}^2 \mid Y, Z\right]. \qquad (4.2)
\end{aligned}
$$

In (4.2) we take into account the fact that $h(Y, Z) \equiv E(X \mid Y) - E(X \mid Y, Z)$ is Y, Z measurable and behaves as a constant in conditional expectation with respect to Y, Z. Then we have

$$
\begin{aligned}
&[E(X \mid Y) - E(X \mid Y, Z)]\, E\left[(X - E(X \mid Y)) \mid Y, Z\right] \\
&= [E(X \mid Y) - E(X \mid Y, Z)]\, [E(X \mid Y, Z) - E[E(X \mid Y) \mid Y, Z]] \\
&= [E(X \mid Y) - E(X \mid Y, Z)]\, [E(X \mid Y, Z) - E(X \mid Y)] \\
&= -\, [E(X \mid Y) - E(X \mid Y, Z)]^2 \qquad (4.3)
\end{aligned}
$$

Therefore, taking into account (4.3) in (4.2)

$$
\begin{aligned}
&E\left[(X - E(X \mid Y, Z))^2 \mid Y, Z\right] \\
&= E[(X - E(X \mid Y))^2 \mid Y, Z] \\
&- 2\,[E(X \mid Y) - E(X \mid Y, Z)]^2 \\
&+ E\left[(E(X \mid Y) - E(X \mid Y, Z))^2 \mid Y, Z\right] \qquad (4.4)
\end{aligned}
$$

Applying the expected value operator to both sides of (4.4) we have

$$
\begin{aligned}
E[X - E(X \mid Y, Z)]^2 &= E[X - E(X \mid Y)]^2 \\
&- 2E\left\{[E(X \mid Y) - E(X \mid Y, Z)]^2\right\} \\
&+ E\left\{E\left[\left(E(X \mid Y) - E(X \mid Y, Z)^2\right) \mid Y, Z\right]\right\} \\
&= E[X - E(X \mid Y)]^2 - 2E\left\{[E(X \mid Y) - E(X \mid Y, Z)]^2\right\} \\
&+ E\left\{[E(X \mid Y) - E(X \mid Y, Z)]^2\right\} \\
&= E\left[(X - E(X \mid Y))^2\right] - E\left\{[E(X \mid Y) - E(X \mid Y, Z)]^2\right\}
\end{aligned}
$$

which implies

$$E[X - E(X \mid Y, Z)]^2 \leq E[X - E(X \mid Y)]^2.$$

□

Corollary 4.1. *For any $n \geq 2$, and the sequence of random variables $X_1, X_2, ..., X_n, ...$ it is true that*

$$E\left[X_{n+1} - E(X_{n+1} \mid X_{i_1}, X_{i_2}, ..., X_{i_l})\right]^2 \leq E\left[X_{n+1} - E(X_{n+1} \mid X_{i_1}, X_{i_2}, ..., X_{i_k})\right]^2,$$
$$1 \leq i_1 < i_2 < ... < i_k < i_l \leq n, 1 \leq k < l \leq n.$$

For example,

$$E\left[X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_n)\right]^2 \le E\left[X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_{n-1})\right]^2$$
$$E\left[X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_n)\right]^2 \le E\left[X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_{n-2})\right]^2$$
$$E\left[X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_n)\right]^2 \le E\left[X_{n+1} - E(X_{n+1} \mid X_2, X_3, ..., X_n)\right]^2$$
$$E\left[X_{n+1} - E(X_{n+1} \mid X_2, X_3, X_4\right]^2 \le E\left[X_{n+1} - E(X_{n+1} \mid X_2, X_3)\right]^2$$

etc.

Example 4.1. **(Martingale)** The sequence $X_1, X_2, ..., X_n, ...$ is called a martingale if
$$E(X_{n+1} \mid X_1, X_2, ..., X_n) = X_n.$$

It follows from the Corollary that if $X_1, X_2, ..., X_n, ...$ is a martingale then

$$\begin{aligned} & E[X_{n+1} - X_n]^2 \\ & = E\left[X_{n+1} - E(X_{n+1} \mid X_1, X_2, ..., X_n)\right]^2 \\ & \le E\left[X_{n+1} - E(X_{n+1} \mid X_{i_1}, X_{i_2}, ..., X_{i_k})\right]^2 \\ & 1 \le i_1 < i_2 < ... < i_k \le n, 1 \le k \le n. \end{aligned}$$

Example 4.2. **(Markov chain)** Let $X_1, X_2, ..., X_n, ...$ be a Markov chain, i.e. for an interval A of the real line it is true that

$$P\{X_n \in A \mid X_{i_1}, X_{i_2}, ..., X_{i_k}\} = P\{X_n \in A \mid X_{i_k}\}.$$

Then we have

$$\begin{aligned} E\left[X_n - E(X_n \mid X_{i_l})\right]^2 &= E\left[X_n - E(X_n \mid X_{i_1}, X_{i_2}, ..., X_{i_l})\right]^2 \\ &\le E\left[X_n - E(X_n \mid X_{i_1}, X_{i_2}, ..., X_{i_k})\right]^2 \\ &= E\left[X_n - E(X_{n+1} \mid X_{i_k})\right]^2, \\ & 1 \le i_1 < i_2 < ... < i_k < i_l \le n, 1 \le k < l \le n. \end{aligned} \tag{4.5}$$

Therefore,,

$$E\left[X_{n+1} - E(X_{n+1} \mid X_l)\right]^2 \le E\left[X_{n+1} - E(X_{n+1} \mid X_k)\right]^2, \quad 1 \le k < l \le n.$$

A good illustration of this fact can be given with order statistics.

Example 4.3. **(Order statistics)** Let $X_1, X_2, ..., X_n$ be iid random variables and $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ be the order statistics. The theory of order statistics is well described in David and Nagaraja (2003) ([9]), Arnold, Balakrishnan and Nagaraja (1992) ([5]). It is well known that the order statistics form a Markov chain. Then for $1 \leq k < l \leq n$ one can write

$$E\left[X_{n:n} - E(X_{n:n} \mid X_{l:n})\right]^2 \leq E\left[X_{n:n} - E(X_{n:n} \mid X_{k:n})\right]^2,$$

i.e. $E(X_{n:n} \mid X_{l:n})$ predicts $X_{n:n}$ better than $E(X_{n:n} \mid X_{k:n})$. In particular, for demonstration of this fact consider

$$E[X_{n:n} - E(X_{n:n} \mid X_{n-1:n})]^2 \leq E[X_{n:n} - E(X_{n:n} \mid X_{n-2:n})]^2.$$

It means that $E(X_{n:n} \mid X_{n-1:n})$ is better estimation for $X_{n:n}$ than $E(X_{n:n} \mid X_{n-2:n})$. Let us compute the functions $g_1(x) = E(X_{n:n} \mid X_{n-1:n} = x)$ and $g_2(x) = E(X_{n:n} \mid X_{n-2:n} = x)$. From the joint distribution of $X_{r:n}$ and $X_{s:n}$, $r < s$, we can easily write the conditional pdf's of $X_{n.n} \mid X_{n-1:n}$ and $X_{n.n} \mid X_{n-2:n}$ as

$$f_{n|n-1}(z \mid x) = \frac{f(z)}{1 - F(x)}, x < z$$

$$f_{n|n-2}(z \mid x) = \frac{2(F(z) - F(x))}{(1 - F(x))^2} f(z), x < z,$$

respectively. Then for a $Uniform(0,1)$ distribution, we can write

$$g_1(x) = E(X_{n:n} \mid X_{n-1:n} = x) = \int_x^1 f_{n|n-1}(z \mid x)dz$$

$$= \frac{1+x}{2}, 0 \leq x \leq 1$$

and

$$g_2(x) = E(X_{n:n} \mid X_{n-2:n} = x) = 2\int_x^1 f_{n|n-2}(z \mid x)dz$$

$$= \frac{x+2}{3}, 0 \leq x \leq 1$$

and it is clear that

$$g_1(x) < g_2(x), 0 \leq x \leq 1$$

because

$$\frac{x+2}{3} = \frac{1+x}{2} + \frac{1-x}{6}, 0 \leq x \leq 1.$$

This means that $g_1(X_{n-1:n}) = E(X_{n:n} \mid X_{n-1:n}) > g_2(X_{n-2:n}) = E(X_{n:n} \mid X_{n-2:n})$ almost sure, hence $E(X_{n:n} \mid X_{n-1:n})$ is better than $E(X_{n:n} \mid X_{n-2:n})$ as a predictor of $X_{n:n}$. For an exponential distribution $F(x) = 1 - \exp(-x), x \geq 0$

it can be easily verify that

$$\begin{aligned} g_1(x) &= E(X_{n:n} \mid X_{n-1:n} = x) \\ &= \frac{1}{1-F(x)} \int_x^\infty z f_{n|n-1}(z \mid x) dz \\ &= \frac{1}{1-F(x)} \int_x^\infty z f(z) dz \\ &= e^x \int_x^\infty z e^{-z} dz = e^x e^{-x}(x+1) = x+1 \end{aligned}$$

i.e. $g_1(x) > g_2(x), x \geq 0$ and again $E(X_{n:n} \mid X_{n-1:n})$ is better than $E(X_{n:n} \mid X_{n-2:n})$ as a predictor of $X_{n:n}$.

Example 4.4. (Record Values) Let $X_1, X_2, ..., X_n, ...$ be a sequence of independent identically distributed (i.i.d.) r.v.'s with continuous d.f. F ; $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ be the order statistics of $X_1, X_2, ..., X_n$. The random variable X_K is called a (upper) record value of the sequence $\{X_n, n \geq 1\}$ if $X_K > \max\{X_1, X_2, ..., X_{K-1}\}$. By convention X_1 is record value. Denote by $\{U(n), n > 1\}$ the sequence of record times:

$$U(n) = \min\left\{j : j > U(n-1), X_j > \ X_{U(n-1)}\right\}, n > 1 \quad \text{with } U(1) = 1.$$

$X_{U(n)}$ is called n th upper record value. Developments on records have been reviewed by many authors including Nevzorov (1988) ([12]), Nagaraja (1988) ([10]), Arnold and Balakrishnan (1989) ([4]), Arnold, Balakrishnan and Nagaraja (1998) ([6]), Ahsanullah (1995) ([1]). The properties of records values of iid random variables have been extensively studied in the literature. Many properties of records can be expressed in terms of the functions $R(x) = -\log \bar{F}(x)$ where $\bar{F}(x) = 1 - F(x)$ and $0 < \bar{F}(x) < 1$. It is well known that, the sequence of record values $X_{U(1)}, X_{U(2)}, ..., X_{U(n)}, ...$ form a Markov chain. From the corollary and (4.5) we have

$$\begin{aligned} &E\left[X_{U(n)} - E(X_{U(n)} \mid X_{U(1)}, X_{U(2)}, ..., X_{U(n-1)})\right]^2 \\ &\leq E\left[X_{U(n)} - E(X_{U(n)} \mid X_{U(1)}, X_{U(2)}, ..., X_{U(n-2)})\right]^2 \end{aligned} \tag{4.6}$$

By Markov property

$$\begin{aligned} E\left(X_{U(n)} \mid X_{U(1)}, X_{U(2)}, ..., X_{U(n-1)}\right) &= E\left(X_{U(n)} \mid X_{U(n-1)}\right) \\ E\left(X_{U(n)} \mid X_{U(1)}, X_{U(2)}, ..., X_{U(n-2)}\right) &= E\left(X_{U(n)} \mid X_{U(n-2)}\right). \end{aligned} \tag{4.7}$$

Then from (4.6) and (4.7) for any $n > 2$ we have

$$E\left[X_{U(n)} - E(X_{U(n)} \mid X_{U(n-1)})\right]^2 \leq E\left[X_{U(n)} - E(X_{U(n)} \mid X_{U(n-2)})\right]^2, \tag{4.8}$$

i.e. $E(X_{U(n)} \mid X_{U(n-1)})$ is better than $E(X_{U(n)} \mid X_{U(n-2)})$ as a predictor of $X_{U(n)}$. It is clear that, (4.8) can be extended as

$$E\left[X_{U(n)} - E(X_{U(n)} \mid X_{U(l)})\right]^2 \leq E\left[X_{U(n)} - E(X_{U(n)} \mid X_{U(k)})\right]^2, \quad 2 < k < l < n. \tag{4.9}$$

It is possible to extend the list of examples to the areas where the prediction of random variables with conditional expectations is the subject.

5. CONCLUSION

This paper investigates the new properties of conditional expectation with respect to a sigma-algebra generated by other random variables. The conditional expectations of the random variable with respect to a sigma-algebra generated by the random variable is its best predictor in the sense of least square distance. Some important inequalities concerning the predictions of random variables are proved. These inequalities can find important applications in many areas such as financial mathematics, actuarial sciences, and reliability engineering. An application of the main inequality having interesting consequences in per-share stock is presented. Considering conditional expectations as random variables, we study also the dependence properties of and copulas of these random variables. Some examples with ordered random variables and martingales are provided.

ACKNOWLEDGMENTS

I thank the anonymous referee and an editor for their useful comments which resulted in improvements in the presentation of this paper. I thank also Professor Majid Asadi who suggested an extension of Theorem 2.1 mentioned in Remark 2.1.

REFERENCES

[1] Ahsanullah, M. (1995). *Record Statistics*, Nova Science Publishers, Inc., New York.

[2] Ahsanullah, M. and Nevzorov, V.B . (2001). *Ordered Random Variables*, Nova Science Publishers, Huntington, NY.

[3] Ahsanullah, M. and Nevzorov, V.B . (2005). *Order Statistics.Examples and Exercises*, Nova Science Publishers, Huntington, NY.

[4] Arnold, B.C. and Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics*, Lecture Notes in Statistics, 53, Springer, New York.

[5] Arnold, B.C.; Balakrishnan, N. and Nagaraja, H.N. (1992). *First Course in Order Statistics*, Wiley, New York.

[6] Arnold, B.C.; Balakrishnan, N. and Nagaraja, H.N. (1998). *Records*, Wiley, New York.

[7] Arnold, B.C.; Becker A.; Gather, U. and Zahedi, H. (1984). On the Markov property of order statistics, *Journal of Statistical Planning and Inference*, **9**, 147–154.

[8] Borovkov, A.A. (1998). *Mathematical Statistics*, Gordon and Breach Science Publishers.

[9] David, H.A. and Nagaraja, H.N. (2003). *Order Statistics*, Third ed., Wiley, New Jersey.

[10] Nagaraja, H.N. (1988). Record values and related statistics - a review, *Communications in Statistics - Theory and Methods*, **17**, 2223–2238.

[11] Nelsen, R. B. (2006). *An Introduction To Copulas*, Second Edition. Springer.

[12] Nevzorov, V.B. (1988). Records, *Theory of Probability and its Applications*, **32(2)**, 201–228.

[13] Ross, S. (2002). *A first course in Probability*, Prentice Hall.