Lecture 8 / Week 5

OUTLINE

1) An example
2) Best Forecast Scheme
3) Conditioning on Increasing σ-Algebras
4) Distributions of Transformations of Random Vectors
5) Noteworthy Distributions

Example Suppose we have two random variables X, Y. We know that X has the exponential (λ) distribution with the density function (absolutely continuous)

$$f_X(x) = \lambda.e^{-\lambda.x}\mathbf{1}_{[0,\infty)}(x)$$

We also know that the conditional probability distribution is discrete: $(Y|X = x) \sim Poisson(x)$ Note $\lambda = x$.

$$P(Y = y|X = x) = \frac{e^{-x}.x^y}{y!} \qquad y = 1, 2....$$

Suppose we want to compute $E(X.Y) = E(g(X, Y))$. The problem is the function g has one discrete and one abs. continuous component, and we do not know how to compute the expectation in such a case. BUT, we can exploit the properties of the conditional expectation, i.e

$$E(X.Y) = E(E(X.Y|X)) = E(X.E(Y|X))$$

But we know that the expected value of a variable with poisson distribution is $E(X) = \lambda$. Consequently, $E(Y|X = x) = x$, or $E(Y|X) = X$, but then

$$E(X.E(Y|X)) = E(X.X) = E(X^2)$$

Then using the formula

$$var(X) = E(X^2) - [E(X)]^2$$

and recalling that the expected value and the variance of a variable with exponential distribution; $E(X) = \frac{1}{\lambda}, var(X) = \frac{1}{\lambda^2}$, respectively, we have

$$E(X^2) = var(X) + [E(X)]^2 = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}. \quad \text{QED.}$$

This example is a good illustration where we have to exploit the properties of conditional expectation.

Best Forecast Scheme=Conditional Expectation

The conditional expectation $E(Y|X)$ has a special interpretation once the second moment of the variable Y has finite moment, formally $E(Y^2) < \infty$. Namely, it is the best forecast of Y based on X; i.e suppose we have a random variable Y and a random vector X and we want to predict Y as a function of X. Then we claim that, which will prove in a second, $g(X)$ is the best predictor of Y, in the sense that it minimizes the prediction error. We need the following definition to clarify the problem

Definition We define the **mean square error of prediction** (**MSEP**) as follows $E[(Y - g(X))^2]$. In words, this is the expected error we are trying to minimize, naturally it is the expectation of the square of difference between the actual random value(Y) and our prediction(g(X)). Note that it is squared so that positive and negative errors cannot cancel out.

Theorem The funciton g that minimizes MSEP is $g(X) = E(Y|X)$. Formally, for every function $g(X)$

$$E((Y - g(X))^2) \geq E((Y - E(Y|X))^2)$$

Proof We take $E((Y - g(X))^2)$ and do the add/substract trick, i.e

$$E((Y - E(Y|X) + E(Y|X) - g(X))^2)$$

we group the terms and take the square

$$E((Y - g(X))^2) = E[((Y - E(Y|X)) + (E(Y|X) - g(X)))^2] =$$

$= E[((Y-E(Y|X))^2]+E[(E(Y|X)-g(X))^2]+2.E[(Y-E(Y|X)).(E(Y|X)-g(X))]$ (*)

Then we show that

$$\begin{aligned} E[(Y - E(Y|X)).(E(Y|X) - g(X))] &= 0 \\ E[E((Y - E(Y|X)).(E(Y|X) - g(X))|X)] &= 0 \end{aligned}$$

In the second line we used the conditional expectation property, i.e. the expectation of the conditional expectation is the expectation itself, since the second term of the multiplication is fixed at X, behaves like a constant, we can take it out,

$$E[(E(Y|X) - g(X)).E(Y - E(Y|X)|X)]$$

but then

$$E(Y - E(Y|X)|X) = E(Y|X) - E(Y|X) = 0$$

Since in (*), the only term left that is dependent on $g(X)$ is $E[(E(Y|X) - g(X))^2]$, since it is a square, we can minimize the error by setting

$$E(Y|X) = g(X). \qquad QED.$$

The previous result has important implication in econometrics and particaularly it is fundamental in regression analysis.

Conditioning on Increasing σ-Algebras

We might want to make predictions in more general settings. Assume that we have the following sequence of random variables

$$\{Y_t\}_{t=-\infty}^{\infty}$$

Then we might want to know what the best prediction Y_t is based on the Y's at times t-1,t-2.....t-m.

$$E(Y_t|Y_{t-1}, Y_{t-2}......Y_{t-m})$$

as m goes larger($m \rightarrow \infty$) the σ-algebra (the information set) becomes larger as well. We interested in answering questions such as; what happens to the conditional expectation as the sequence of σ-algebra's goes larger, i.e

$$What\ is\ \lim_{m\rightarrow\infty} E(Y_t|Y_{t-1}, Y_{t-2}......Y_{t-m})?$$

The answer lies in the following theorem.

Theorem Let $\mathcal{F}_n$ be an increasing sequence of σ-algebras. (i.e. $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$). Let $\mathcal{F}_\infty = \vee_{n=1}^{\infty}\mathcal{F}_n$, which is the σ-algebra that contains the union of σ-algebras (since the union itself is not necessarily a σ-algebra.) Then

$$E(Y|\mathcal{F}_n)_{n\rightarrow\infty} \rightarrow E(Y|\mathcal{F}_\infty) \quad \textbf{a.s}$$

Proof It is complicated and thus omitted.

As mentioned above we are interested in

$$E(Y_t|Y_{t-1}, Y_{t-2}......)$$

but now following the theorem we know that

$$E(Y_t|Y_{t-1}, Y_{t-2}......Y_{t-m}) \approx E(Y_t|Y_{t-1}, Y_{t-2}......)$$

if we have m large enough, in other words if we can collect as much past data as possible (m large depends on the data and application, e.g whether data is quarterly, weekly, or so) than we can be almost sure that we have all the past information about the random variable and the expectation conditioned (based on this information set) on this past data is going to be the best prediction of Y_t.

Distributions of Transformations of Random Vectors

We have the random vector X and the function g s.t $Y = g(X)$. We want to know how to compute the distribution of Y once we know the distribution of X. We will again analyse two different cases; discrete and absolutely continuous.

Case 1 We have a **discrete random vector** X :

$$p_X(x) = P(X = x) \quad S = \{x_1, x_2....\} \text{ at most countable}$$

Then we will also have a discrete Y s.t $Y = g(X)$ and $S_y = \{y_1, y_2.....\}$ at most countable. Note that in fact we have a vector X so the correct notation should be

$$p_{X_1......X_k}$$

but we slightly abuse the notation. Then we will also have $g(x_i)$; $\{g(x_1), g(x_2).....g(x_k)\}$, but note that this set can contain less elements than k distinct functions, if some functions have the same value, in fact it only includes functions g with distinct values. Then

$$p_Y(y_i) = P(Y = y_i) = P(X \in g^{-1}(y_i)) = \sum_{g(x_j)=y_i} p_X(x_j)$$

We can better see this with an example:

Example We have the random vector $X = (X_1, X_2)$ where X_1, X_2 have independent Poisson(λ) distributions. We also have

$$\begin{aligned} g(x_1, x_2) &= x_1 + x_2 \\ Y &= X_1 + X_2 \\ S_x &= \{(0,0), (0,1), (1,0).....\} \\ s_1 &= 0, 1, 2........ \\ s_2 &= 0, 1, 2........ \\ S_y &= \{0, 1, 2......\} \end{aligned}$$

We are trying to calculate

$$P(Y = y) = P(X_1 + X_2 = y) \qquad y = 0, 1, 2.....$$

Using the previous formula and the indepence of two components

$$\sum_{x_1+x_2=y} P(X_1 = x_1, X_2 = x_2) =^{ind.} \sum_{x_1+x_2=y} P(X_1 = x_1).P(X_2 = x_2)$$

we use the poisson distribution

$$\begin{aligned} \sum_{x_1+x_2=y} P(X_1 &= x_1).P(X_2 = x_2) = \sum_{x_1+x_2=y} \frac{e^{-\lambda}.\lambda^{x_1}}{x_1!}.\frac{e^{-\lambda}.\lambda^{x_2}}{x_2!} = \\ P(Y &= y) = \frac{e^{-2\lambda}.(2\lambda)^y}{y!} \quad y = 0, 1, 2... \end{aligned}$$

Exercise Make the above calculation

$$\sum_{x_1+x_2=y} \frac{e^{-\lambda}.\lambda^{x_1}}{x_1!}.\frac{e^{-\lambda}.\lambda^{x_2}}{x_2!} = e^{-2\lambda}.\lambda^y.\sum_{x_1+x_2=y} \frac{1}{x_1!.x_2!}$$

Hint multiply divide by $y!$ and use $(1+1)^y = \sum \binom{y}{x} 1^x.1^{y-x}$.

An important point to note from this example is that the sum of two independent poisson variables ended up to have a poisson distribution with the parameter $(2.\lambda)$.

Case 2 X has **absolutely continuous distribution:** We know that then we have

$$P(X \in B) = \int_B f_X(x)dx$$

Suppose we have $Y = g(X)$. Does the absoulute continuity of X imply the same for Y? The answer is no in general. Here we have a counterexample;

Example $X \sim N(0,1)$,so it has standard normal distribution and Y is the following function of X :

$$\begin{aligned} Y &= \left\{ \begin{array}{ll} 0 & X \geq 0 \\ 1 & X < 0 \end{array} \right\} \\ Y &= \mathbf{1}_{(-\infty,0)}(X) \end{aligned}$$

but Y is a discrete random variable. In fact Y has Bernoulli($\frac{1}{2}$) distribution.

$$\begin{aligned} P(Y &= 0) = P(X > 0) = \frac{1}{2} \\ P(Y &= 1) = P(X \leq 0) = \frac{1}{2} \end{aligned}$$

In general, once Y has a absolutely continuous distribution we have

$$\begin{aligned} P(Y &\in B) = P(g(X) \in B) = P(X \in g^{-1}(B)) = \\ &= \int_{g^{-1}(B)} f_X(x)dx \\ P(Y &\leq y) = \int_{g^{-1}(-\infty,y]} f_X(x)dx \end{aligned}$$

Still we have to impose some conditions to be able to find the density of Y directly from the density of X.

Theorem Let X be a k-dimensional random vector with absolutely continuous distribution and the density function $f_X(x)$. Let $g : \mathbb{R}^k \to \mathbb{R}^k$ be a

one to one function such that its **inverse** function $g^{-1}$is **differentiable**. Let $Y = g(X)$. Then Y has absolutely continuous distribution with density function

$$\begin{aligned}
f_Y(y) &= f_X(g^{-1}(y)).|\det J(y)| \text{ on the set} \\
\{y &: y = g(x) \text{ with } f_X(x) > 0\} \\
where\ J_{ij}(y) &= \frac{\partial g_i^{-1}}{\partial y_j} \text{ (Jacobian)}
\end{aligned}$$

The following example illustrates such a case

Example Suppose X_1, X_2 independent $N(0,1)$, i.e.

$$f_{X_{1,2}}(x) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}x^2}$$

Also suppose that we have (Y_1, Y_2) , so 1-1 requirement satisfied and

$$\begin{aligned}
Y_1 &= X_1 + X_2 \Rightarrow X_1 = \frac{Y_1 + Y_2}{2} \\
Y_2 &= X_1 - X_2 \Rightarrow X_2 = \frac{Y_1 - Y_2}{2}
\end{aligned}$$

so the functions $g_{1,2}$ are invertible (1-1)

$$\begin{aligned}
X_1 &= g_1^{-1}(Y_1, Y_2) = \frac{Y_1 + Y_2}{2} \\
X_2 &= g_2^{-1}(Y_1, Y_2) = \frac{Y_1 - Y_2}{2}
\end{aligned}$$

so they are also differentiable (condition). Then we can compute the Jacobian

$$\begin{aligned}
J &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \\
|\det(J)| &= \frac{1}{2}
\end{aligned}$$

Applying the formula and exploting the indepence we have

$$\begin{aligned}
f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1}(\frac{Y_1 + Y_2}{2}).f_{X_2}(\frac{Y_1 - Y_2}{2}).\frac{1}{2} \\
&= \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}(\frac{y_1+y_2}{2})^2}.\frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}(\frac{y_1-y_2}{2})^2}.\frac{1}{2} = \\
&= \frac{1}{\sqrt{2\pi}.\sqrt{2}}.e^{-\frac{1}{2}\frac{y_1^2}{2}}.\frac{1}{\sqrt{2\pi}.\sqrt{2}}.e^{-\frac{1}{2}\frac{y_2^2}{2}}
\end{aligned}$$

So we have found that Y_1 and Y_2 are independent and have the normal distribution $N(0, 2)$.

Noteworthy Distributions

1. **Normal Distribution**: It is denoted by $N(\mu, \sigma^2)$

$$\begin{aligned} E(X) &= \mu \\ var(X) &= \sigma^2 \\ f_X(x) &= \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}\frac{(x-\mu)}{\sigma^2}^2} \end{aligned}$$

2. **Standard Normal Distribution**: It is denoted by $N(0, 1)$. It is a special case of normal distribution

$$\begin{aligned} E(X) &= 0 \\ var(X) &= 1 \\ E(X^4) &= 3 \\ f_X(x) &= \frac{1}{\sqrt{2\pi}}.e^{-\frac{1}{2}x^2} \end{aligned}$$

Exercise If X$\sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2.\sigma^2)$

Proof It directly folllows from the properties of expectation and variance.

$$\begin{aligned} E(X) &= \mu \\ E(Y) &= a.E(X) + E(b) = a\mu + b \\ var(X) &= \sigma^2 \\ var(Y) &= a^2 var(X) = a^2.\sigma^2 \quad \text{var(b)=0} \end{aligned}$$

3. **Chi-Square Distribution**: It is denoted by $\mathcal{X}_n^2$. Let $X_1, X_2....X_n$ be independent identically distributed random variables with standard normal distributions and set $Y = X_1^2 + X_2^2.... + X_n^2$. This new random variable Y has chi-square distribution with n *degrees of freedom*.

$$\begin{aligned} f_Y(y) &= \frac{1}{\Gamma(\frac{n}{2}).2^{\frac{n}{2}}}, y^{\frac{n}{2}-1}.e^{-\frac{y}{2}}.\mathbf{1}_{(0,\infty)}(y) \\ E(Y) &= n.E(X_1^2) = n \\ var(Y) &= n.var(X_1^2) = n.[E(X_1^4) - (E(X_1^2))^2] = n(3-1) = 2n \end{aligned}$$

where $\Gamma(\alpha)$ is called gamma function.

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}.e^{-x}dx \quad \text{for } \alpha > 0$$

Exercise Show that $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

Proof Since $\Gamma(\alpha-1) = \int_0^\infty x^{\alpha-2}.e^{-x}dx$ for $\alpha > 0$, we integrate by parts $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}.e^{-x}dx$=$[-e^{-x}.x^{\alpha-1}]_0^\infty$+$\int_0^\infty(\alpha-1)x^{\alpha-2}.e^{-x}dx = (\alpha-1)\Gamma(\alpha-1)$.

Exercise $\Gamma(1) = 1$.

Proof Since $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}.e^{-x}dx$ for $\alpha > 0$, then $\Gamma(1) = \int_0^\infty x^0.e^{-x}dx = 1$

Exercise $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Note that $\Gamma(n) = (n-1)!$, this function is a generalization of the factorials where n can be real numbers not just integers.

4. **t Distribution** X, Y are independent X$\sim N(0,1)$ and $Y \sim \mathcal{X}_n^2$.

$$T = \frac{X}{\sqrt{\frac{Y}{n}}} \quad t-distribution$$

5. **F(Fisher) Distribution** $X \sim \mathcal{X}_m^2$, $Y \sim \mathcal{X}_n^2$

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}} \quad F_{m,n}$$