Core course for *App.AI*, *Bioinfo.*, *Data Sci.&Eng.*, *Dec.Analytics*, *Q.Fin.*, *Risk Mgmt.* and *Stat.* Majors:

STAT2601A Probability and Statistics I (2023-2024 First Semester)

Chapter 8: Conditional Distributions

8.1 Conditional Distributions

Often times when two random variables, X and Y, are observed, the values of the two variables are related. For example, suppose that

$$X = \text{a person's height}, \qquad Y = \text{the same person's weight}.$$

Surely we would think it more likely that $Y > 200$ pounds if we were told that $X = 182$ cm than if we were told that $X = 104$ cm.

The knowledge about the value of X gives us some information about the value of Y even though it does not reveal the exactly value of Y.

Recall that for any two events E and F, the conditional probability of E given F is defined by

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)} \qquad \text{provided that } \Pr(F) > 0.$$

This leads to the following definition.

Definition 8.1.
Let (X, Y) be a discrete bivariate random vector with joint pmf $p(x, y)$ and marginal pmfs $p_X(x)$ and $p_Y(y)$. For any x such that $p_X(x) = \Pr(X = x) > 0$, the *conditional pmf* of Y *given that* $X = x$ is the function of y denoted by $p_{Y|X}(y|x)$ and is defined by

$$p_{Y|X}(y|x) = \Pr(Y = y|X = x) = \frac{\Pr(Y = y, X = x)}{\Pr(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

On the other hand, for any y such that $p_Y(y) = \Pr(Y = y) > 0$, the *conditional pmf* of X *given that* $Y = y$ is the function of x denoted by $p_{X|Y}(x|y)$ and is defined by

$$p_{X|Y}(x|y) = \Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

◀

Example 8.1.

Referring to **Example 7.1.** in **Chapter 7** that we randomly draw 3 balls from an urn with 3 red balls, 4 white balls, 5 blue balls. Let X be number of red balls, Y be the number of white balls in the sample. The joint pmf of (X, Y) is given by the following table.

Value of X	Value of Y				Total
	0	1	2	3	
0	0.0454	0.1818	0.1364	0.0182	0.3818
1	0.1364	0.2727	0.0818	0	0.4909
2	0.0682	0.0545	0	0	0.1227
3	0.0045	0	0	0	0.0045
Total	0.2545	0.5091	0.2182	0.0182	1.0000

Dividing all the entries by the row totals we obtain the conditional pmfs of $Y|(X = x)$.

Value of X	Value of Y				Total
	0	1	2	3	
0	0.1190	0.4762	0.3571	0.0476	1
1	0.2778	0.5556	0.1667	0	1
2	0.5556	0.4444	0	0	1
3	1	0	0	0	1

Each row represents a conditional pmf of $Y|(X = x)$. For example, the first row is the conditional pmf of Y given that $X = 0$, the second row is the conditional pmf of Y given that $X = 1$, etc. From these conditional pmfs we can see how our uncertainty on the value of Y is affected by our knowledge on the value of X.

Similarly, dividing all the entries in the joint pmf table by the column totals gives the conditional pmf of $X|(Y = y)$ which is shown in the following table.

Value of X	Value of Y			
	0	1	2	3
0	0.1786	0.3571	0.6250	1
1	0.5357	0.5357	0.3750	0
2	0.2679	0.1071	0	0
3	0.0179	0	0	0
Total	1	1	1	1

Example 8.2.

Let $X \sim \text{Po}(\lambda_1)$ and $Y \sim \text{Po}(\lambda_2)$ be two independent Poisson random variables. If it is known that $X + Y = n > 0$, what will be the conditional distribution of X?

Obviously, the possible values of X given $X + Y = n$ should be from 0 to n.

First of all, the random variable $X + Y$ is distributed as $\text{Po}(\lambda_1 + \lambda_2)$ as the moment generating function of $X + Y$ is

$$M_{X+Y}(t) = M_X(t)M_Y(t) = e^{\lambda_1(e^t-1)} \times e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}.$$

Now consider

$$\begin{aligned}
\Pr(X = k|X + Y = n) &= \frac{\Pr(X = k, X + Y = n)}{\Pr(X + Y = n)} \\
&= \frac{\Pr(X = k, Y = n - k)}{\Pr(X + Y = n)} \\
&= \frac{\Pr(X = k)\Pr(Y = n - k)}{\Pr(X + Y = n)} \\
&= \left[\frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right] \bigg/ \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!} \qquad \because X + Y \sim \text{Po}(\lambda_1 + \lambda_2) \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}, \qquad k = 0, 1, \ldots, n.
\end{aligned}$$

Hence, $X|(X + Y = n) \sim \text{B}\left(n, \dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

★

Remarks

1. If X is independent of Y, then the conditional pmf of X given $Y = y$ will becomes

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x) \qquad \text{for all } y.$$

Similarly,

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{p_X(x)p_Y(y)}{p_X(x)} = p_Y(y) \qquad \text{for all } x.$$

Hence, the knowledge of the value of one variable does not affect the uncertainty on the value of another variable, i.e., knowledge of one variable gives no information on the other variable if they are independent.

2. For continuous random variables, the conditional distributions are defined in an analogous way.

Definition 8.2.
Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$.
The *conditional pdf* of Y *given that* $X = x$ is defined by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \qquad \text{provided that } f_X(x) > 0.$$

The *conditional pdf* of X *given that* $Y = y$ is defined by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \qquad \text{provided that } f_Y(y) > 0.$$

◀

3. The conditional distribution of Y given $X = x$ is possibly a different probability distribution for each value of x. Thus we actually have a family of probability distributions for Y, one for each x. When we wish to describe this entire family, we will use the phrase "the distribution of $Y|(X = x)$".

8.2 Properties of Conditional Distributions

The conditional pmf/pdf satisfies all the properties of a pmf/pdf and describes the probabilistic behaviour of a random variable given the value of another random variable. Hence we can have the followings definitions.

Definition 8.3.
Conditional Distribution Function of Y given $X = x$:

$$F_{Y|X}(y|x) = \Pr(Y \le y|X = x) = \begin{cases} \sum_{i \le y} p_{Y|X}(i|x), & \text{for discrete case;} \\ \int_{-\infty}^{y} f_{Y|X}(t|x)\mathrm{d}t, & \text{for continuous case.} \end{cases}$$

Conditional Expectation of $g(Y)$ given $X = x$:

$$\mathrm{E}\left[g(Y)|X = x\right] = \begin{cases} \sum_{i} g(i)p_{Y|X}(i|x), & \text{for discrete case;} \\ \int_{-\infty}^{\infty} g(t)f_{Y|X}(t|x)\mathrm{d}t, & \text{for continuous case.} \end{cases}$$

Conditional Mean of Y given $X = x$:

$$\mathrm{E}(Y|X = x)$$

Conditional Variance of Y given $X = x$:

$$\begin{aligned} \mathrm{Var}(Y|X = x) &= \mathrm{E}\left\{[Y - \mathrm{E}(Y|X = x)]^2 \,|X = x\right\} \\ &= \mathrm{E}(Y^2|X = x) - [\mathrm{E}(Y|X = x)]^2 . \end{aligned}$$

◀

Example 8.3.

Suppose that the joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y}, & x > 0, y > 0; \\ 0, & \text{otherwise.} \end{cases}$$

The marginal pdf of Y is

$$f_Y(y) = \int_0^\infty \frac{e^{-x/y}e^{-y}}{y} \mathrm{d}x = e^{-y} \left[e^{-x/y}\right]_\infty^0 = e^{-y}, \qquad y > 0.$$

Hence, $Y \sim \mathrm{Exp}(1)$.

The conditional pdf of $X|(Y=y)$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{y}e^{-x/y}, \qquad x > 0.$$

Hence, the conditional distribution of X given $Y = y$ is exponential with parameter $\lambda = 1/y$, or we may write

$$X|Y \sim \mathrm{Exp}(Y^{-1}), \qquad \text{or} \qquad X|(Y=y) \sim \mathrm{Exp}(y^{-1}).$$

The conditional distribution function of $X|(Y=y)$ is

$$F_{X|Y}(x|y) = \begin{cases} 0, & x \le 0; \\ 1 - e^{-x/y}, & x > 0. \end{cases}$$

Also, the conditional mean and variance can be determined easily as

$$\mathrm{E}(X|Y) = Y, \qquad \mathrm{Var}(X|Y) = Y^2.$$

Therefore $\mathrm{E}(X|Y)$ and $\mathrm{Var}(X|Y)$ are random variables.

Thought Question:

What is $\mathrm{E}\left[\mathrm{E}(X|Y)\right]$?

★

8.3 Computing Expectations by Conditioning

Two important and useful formulae of conditional expectation are given below.

Theorem 8.1. (Law of Total Expectation, or Adam's Law)
If X and Y are two random variables, then for any function u,

$$\mathrm{E}[u(X)] = \mathrm{E}\left\{\mathrm{E}[u(X)|Y]\right\}.$$

Proof. The following proof is for continuous version. The proof for discrete version is similar.

$$\begin{aligned}
\mathrm{E}\left\{\mathrm{E}[u(X)|Y]\right\} &= \int_{-\infty}^{\infty} \mathrm{E}[u(X)|Y=y] f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \left\{\int_{-\infty}^{\infty} u(x) f_{X|Y}(x|y)\mathrm{d}x\right\} f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x) f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \mathrm{E}[u(X)].
\end{aligned}$$

□

In particular, when u is the identity function, we have

$$\mathrm{E}(X) = \mathrm{E}\left[\mathrm{E}(X|Y)\right].$$

▲

Theorem 8.2. (Law of Total Variance, or Eve's Law)
If X and Y are two random variables, then

$$\mathrm{Var}(X) = \mathrm{E}\left[\mathrm{Var}(X|Y)\right] + \mathrm{Var}\left[\mathrm{E}(X|Y)\right].$$

Proof. Consider

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Var}(X|Y)\right] &= \mathrm{E}\left\{\mathrm{E}(X^2|Y) - [\mathrm{E}(X|Y)]^2\right\} \\
&= \mathrm{E}(X^2) - \mathrm{E}\left\{[\mathrm{E}(X|Y)]^2\right\}.
\end{aligned}$$

Then consider,

$$\begin{aligned}
\mathrm{Var}\left[\mathrm{E}(X|Y)\right] &= \mathrm{E}\left\{[\mathrm{E}(X|Y)]^2\right\} - \left\{\mathrm{E}\left[\mathrm{E}(X|Y)\right]\right\}^2 \\
&= \mathrm{E}(X^2) - \mathrm{E}\left[\mathrm{Var}(X|Y)\right] - [\mathrm{E}(X)]^2 \\
&= \mathrm{Var}(X) - \mathrm{E}\left[\mathrm{Var}(X|Y)\right].
\end{aligned}$$

Hence,

$$\mathrm{Var}(X) = \mathrm{E}\left[\mathrm{Var}(X|Y)\right] + \mathrm{Var}\left[\mathrm{E}(X|Y)\right].$$

□

▲

Example 8.4.

In **Example 8.3.**, the marginal pdf of Y is

$$f_Y(y) = e^{-y}, \qquad y > 0.$$

It can be easily verified that $\mathrm{E}(Y) = 1$ and $\mathrm{Var}(Y) = 1$.

Also recall that

$$X|Y \sim \mathrm{Exp}(Y^{-1}),$$

so the conditional mean and variance are respectively,

$$\mathrm{E}(X|Y) = Y, \qquad \mathrm{Var}(X|Y) = Y^2.$$

Therefore,

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}\left[\mathrm{E}(X|Y)\right] = \mathrm{E}(Y) = 1, \\
\mathrm{E}\left[\mathrm{Var}(X|Y)\right] &= \mathrm{E}(Y^2) = \mathrm{Var}(Y) + \left[\mathrm{E}(Y)\right]^2 = 1 + 1^2 = 2, \\
\mathrm{Var}\left[\mathrm{E}(X|Y)\right] &= \mathrm{Var}(Y) = 1, \\
\mathrm{Var}(X) &= \mathrm{E}\left[\mathrm{Var}(X|Y)\right] + \mathrm{Var}\left[\mathrm{E}(X|Y)\right] = 2 + 1 = 3.
\end{aligned}$$

Note that directly calculation of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ from $f(x, y)$ may be difficult as there is no closed form expression for the marginal pdf of X:

$$f_X(x) = \int_0^\infty \frac{e^{-x/y} e^{-y}}{y} \mathrm{d}y.$$

★

Example 8.5.

Suppose we have a binomial random variable X which represents the number of success in n independent Bernoulli experiments. Sometimes the success probability p is unknown. However, we usually have some understanding on the value of p, e.g., we may believe that p is a realization of another random variable P picked uniformly from $(0, 1)$, i.e., $P \sim \mathrm{U}(0, 1)$. Then we have the following *hierarchical model*:

$$P \sim \mathrm{U}(0, 1), \qquad X|P \sim \mathrm{B}(n, P) \text{ or } X|(P = p) \sim \mathrm{B}(n, p).$$

Using the formulae of expectation by conditioning, we have

$$\begin{aligned}
\mathrm{E}(X) &= \mathrm{E}\left[\mathrm{E}(X|P)\right] = \mathrm{E}(nP) = n\mathrm{E}(P) = \frac{n}{2}, \\
\mathrm{Var}(X) &= \mathrm{E}\left[\mathrm{Var}(X|P)\right] + \mathrm{Var}\left[\mathrm{E}(X|P)\right] \\
&= \mathrm{E}\left[nP(1 - P)\right] + \mathrm{Var}\left[nP\right] \\
&= n\mathrm{E}(P) - n\mathrm{E}(P^2) + n^2\mathrm{Var}(P) \\
&= \frac{n}{2} - \frac{n}{3} + \frac{n^2}{12} \\
&= \frac{n(n + 2)}{12}.
\end{aligned}$$

To find the marginal pmf of X, $p_X(x)=\Pr(X=x)$, we can let

$$\mathbf{1}_{\{X=x\}}=\begin{cases}1, & \text{if } X=x;\\ 0, & \text{otherwise.}\end{cases}$$

Then,

$$\begin{aligned}\Pr(X=x) &= \mathrm{E}(\mathbf{1}_{\{X=x\}})\\ &= \mathrm{E}\left[\mathrm{E}(\mathbf{1}_{\{X=x\}}|P)\right]\\ &= \mathrm{E}\left[\Pr(X=x|P)\right]\\ &= \mathrm{E}\left[\binom{n}{x}P^x(1-P)^{n-x}\right]\\ &= \binom{n}{x}\int_0^1 p^x(1-p)^{n-x}(1)\mathrm{d}p \qquad \because f_P(p)=1 \text{ for } 0<p<1\\ &= \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}\\ &= \binom{n}{x}\frac{x!(n-x)!}{(n+1)!}\\ &= \frac{1}{n+1}, \qquad x=0,1,2,\ldots,n.\end{aligned}$$

Hence, X is distributed as discrete uniform distribution with support $\{0,1,2,...,n\}$.

Using the Bayes' theorem, the conditional pdf of P given $X=x$ is given by

$$\begin{aligned}f_{P|X}(p|x) &= \frac{p_{X|P}(x|p)f_P(p)}{p_X(x)}\\ &= \binom{n}{x}p^x(1-p)^{n-x}\times 1\Big/\left(\frac{1}{n+1}\right)\\ &= \frac{(n+1)!}{x!(n-x)!}p^x(1-p)^{n-x}\\ &= \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}p^{(x+1)-1}(1-p)^{(n-x+1)-1}, \qquad 0<p<1.\end{aligned}$$

Therefore, $P|(X=x)\sim \mathrm{Beta}(x+1,n-x+1)$ and

$$\mathrm{E}(P|X=x)=\frac{x+1}{(x+1)+(n-x+1)}=\frac{x+1}{n+2}.$$

In layman's terms, suppose an event happens with an unknown probability P where the values of P are equally likely between 0 and 1, then when x out of n cases of the event were observed, an appropriate estimate of P is

$$\hat{P}=\frac{x+1}{n+2}.$$

This formula is known as the *Laplace's law of succession* in the 18th century by Pierre-Simon Laplace in the course of treating the *sunrise problem* which tried to answer the question "What is the probability that the sun will rise tomorrow?"

★

Example 8.6.

Let $X \sim \text{Geo}(p)$ and $Y \sim \text{Geo}(p)$ be two independent geometric random variables. Find the expected value of the proportion $\dfrac{X}{X+Y}$.

Solution:

Note that $X, Y \in \{1, 2, \ldots\}$. Let $N = X + Y \in \{2, 3, \ldots\}$. The possible values of X given $N = n$ should be from 1 to $n - 1$. Similar to **Example 8.2.**, we first note that $N \sim \text{NB}(2, p)$ as its mgf is

$$M_N(t) = M_{X+Y}(t) = M_X(t)M_Y(t) = \frac{pe^t}{1-(1-p)e^t} \cdot \frac{pe^t}{1-(1-p)e^t} = \left[\frac{pe^t}{1-(1-p)e^t}\right]^2 \text{ for } t < -\ln(1-p).$$

Then, for $k = 1, 2, \ldots, n-1$,

$$\begin{aligned} \Pr(X=k|N=n) &= \Pr(X=k|X+Y=n) = \frac{\Pr(X=k, X+Y=n)}{\Pr(X+Y=n)} = \frac{\Pr(X=k, Y=n-k)}{\Pr(X+Y=n)} \\ &= \frac{\Pr(X=k)\Pr(Y=n-k)}{\Pr(X+Y=n)} = \frac{(1-p)^{k-1}p \cdot (1-p)^{n-k-1}p}{\binom{n-1}{2-1}p^2(1-p)^{n-2}} \qquad \because X+Y \sim \text{NB}(2,p) \\ &= \frac{1}{n-1}. \end{aligned}$$

That is, $X|(N=n) \sim \text{DU}\{1, 2, \ldots, n-1\}$ as a discrete uniform distribution.

The conditional mean of X given $N = n$ is

$$\text{E}(X|N=n) = \frac{1+2+\cdots+(n-1)}{n-1} = \frac{\frac{1}{2}(n-1)(1+n-1)}{n-1} = \frac{n}{2}.$$

Thus, $\text{E}\left(\dfrac{X}{X+Y}\right) = \text{E}\left(\dfrac{X}{N}\right) = \text{E}\left[\text{E}\left(\dfrac{X}{N}\middle| N\right)\right] = \text{E}\left[\dfrac{1}{N}\text{E}(X|N)\right] = \text{E}\left(\dfrac{1}{N}\cdot\dfrac{N}{2}\right) = \dfrac{1}{2}.$

★

Example 8.7. (Prediction of Y from X)

When X and Y are not independent, we can base on the observed value of X to predict the value of the unobserved random variable Y. That is, we may predict the value of Y by $g(X)$ where g is a function chosen in such a way that the mean squared error (MSE) of the prediction, $Q = \text{E}\left\{[Y - g(X)]^2\right\}$, is minimized.

First we conditional on X, consider

$$\begin{aligned} \text{E}\left\{[Y-g(X)]^2|X\right\} &= \text{E}(Y^2|X) - 2g(X)\text{E}(Y|X) + [g(X)]^2 \\ &= \text{Var}(Y|X) + [\text{E}(Y|X)]^2 - 2g(X)\text{E}(Y|X) + [g(X)]^2 \\ &= \text{Var}(Y|X) + [g(X) - \text{E}(Y|X)]^2. \end{aligned}$$

Hence, $Q = \text{E}\left\{\text{E}\left\{[Y-g(X)]^2|X\right\}\right\} = \text{E}[\text{Var}(Y|X)] + \text{E}\left\{[g(X) - \text{E}(Y|X)]^2\right\}.$

Therefore Q is minimized if we choose $g(x) = \text{E}(Y|X=x)$, i.e., the best predictor of Y given the value of X is $g(X) = \text{E}(Y|X)$. The mean squared error of this predictor is

$$\text{E}\left\{[Y - \text{E}(Y|X)]^2\right\} = \text{E}[\text{Var}(Y|X)] = \text{Var}(Y) - \text{Var}[\text{E}(Y|X)] \le \text{Var}(Y).$$

★

Example 8.8.

Two random variables X and Y are said to have a *bivariate normal distribution* if their joint pdf is

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\},$$

for $-\infty<x<\infty$ and $-\infty<y<\infty$, where μ_X and σ_X^2 are the mean and variance of X; μ_Y and σ_Y^2 are the mean and variance of Y; ρ is the correlation coefficient between X and Y. The distribution is denoted as

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}_2\left[\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix},\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right].$$

Consider the marginal pdf of X.

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y \\
&= C(x)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_Y^2}\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\exp\left[\frac{\rho(x-\mu_X)(y-\mu_Y)}{(1-\rho^2)\sigma_X\sigma_Y}\right]\mathrm{d}y \\
&\qquad\qquad \text{where } C(x)=\frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right], \\
&= C(x)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\exp\left[\frac{\rho(x-\mu_X)}{\sigma_X\sqrt{1-\rho^2}}z\right]\mathrm{d}z \qquad \text{by letting } z=\frac{y-\mu_Y}{\sigma_Y\sqrt{1-\rho^2}}, \\
&= C(x)M_Z\left(\frac{\rho(x-\mu_X)}{\sigma_X\sqrt{1-\rho^2}}\right) \qquad \text{where } M_Z(t) \text{ is the mgf of } Z\sim\mathrm{N}(0,1), \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]\exp\left[\frac{1}{2}\frac{\rho^2(x-\mu_X)^2}{\sigma_X^2(1-\rho^2)}\right] \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right], \qquad -\infty<x<\infty.
\end{aligned}$$

Thus, the marginal distribution of X is $\mathrm{N}(\mu_X,\sigma_X^2)$. The conditional pdf of Y given $X=x$ is

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[y-\mu_Y-\frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2\right\},\ -\infty<y<\infty.$$

Hence, $Y|X\sim\mathrm{N}\left(\mu_Y+\dfrac{\rho\sigma_Y}{\sigma_X}(X-\mu_X),\,(1-\rho^2)\sigma_Y^2\right)$.

The best predictor of Y given the value of X is

$$\mathrm{E}(Y|X)=\mu_Y+\frac{\rho\sigma_Y}{\sigma_X}(X-\mu_X)=\left(\mu_Y-\frac{\rho\sigma_Y}{\sigma_X}\mu_X\right)+\frac{\rho\sigma_Y}{\sigma_X}X=\alpha+\beta X.$$

This is called the *linear regression* of Y on X. (Note that $\beta=\frac{\rho\sigma_Y}{\sigma_X}=\frac{\sigma_{XY}}{\sigma_X^2}$ and $\alpha=\mu_Y-\beta\mu_X$.)

★

~ End of Chapter 8 ~