2.5 Conditional distributions

I.Discrete random variables:

$$P(\xi = x_i, \eta = y_j) = p_{ij}, i, j = 1, 2, \cdots.$$

$$\begin{aligned} P(\eta = y_j | \xi = x_i) &= \frac{P(\eta = y_j, \xi = x_i)}{P(\xi = x_i)} \\ &= \frac{p_{ij}}{p_{i\cdot}}, \end{aligned}$$

where $j = 1, 2, \cdots$. This is the conditional distribution of η conditioning on $\xi = x_i$.

Definition

称$P(\eta = y_j|\xi = x_i)$为在$\xi = x_i$的条件下η的条件概率分布列，简称为条件分布，记为$p_{\eta|\xi}(y_j|x_i)$. 称

$$P(\eta \le y|\xi = x_i) = \sum_{j: y_j \le y} p_{\eta|\xi}(y_j|x_i)$$

为在$\xi = x_i$的条件下η的条件分布函数.

从条件分布的定义和ξ, η的独立性的定义可知，ξ, η独立的充分必要条件是对任何$i, j \ge 1$ 有

$$P(\eta = y_j|\xi = x_i) = P(\eta = y_j).$$

Example

在独立重复伯努里试验中，记p为每次试验"成功"的概率，S_n表示第n次成功时的试验次数. 求(1) 在$S_n = t$的条件下，S_{n+1} 的条件概率分布列; (2) 在$S_{n+1} = w$的条件下，S_n的条件概率分布列.

Example

在独立重复伯努里试验中，记p为每次试验"成功"的概率，S_n表示第n次成功时的试验次数. 求(1) 在$S_n = t$的条件下，S_{n+1} 的条件概率分布列; (2) 在$S_{n+1} = w$的条件下，S_n的条件概率分布列.

解 对$t \le w$，事件$\{S_n = t, S_{n+1} = w\}$ 意味着在w次试验中，第t, w次出现"成功"，在第1次到第$t-1$次中出现$n-1$次"成功",其余均出现"失败". 所以

$$
\begin{aligned}
P(S_n = t, S_{n+1} = w) &= p \cdot p \cdot \binom{t-1}{n-1} p^{n-1} q^{w-(n+1)} \\
&= \binom{t-1}{n-1} p^{n+1} q^{w-(n+1)}.
\end{aligned}
$$

而

$$P(S_n = t) = \binom{t-1}{n-1} p^n q^{t-n}.$$

而

$$P(S_n = t) = \binom{t-1}{n-1} p^n q^{t-n}.$$

从而在$S_n = t$的条件下，S_{n+1}的条件概率分布列为

$$P(S_{n+1} = w | S_n = t) = \frac{P(S_n = t, S_{n+1} = w)}{P(S_n = t)} = pq^{w-t-1}.$$

这意味着，在$S_n = t$的条件下，$S_{n+1} - S_n$服从几何分布.

而在$S_{n+1}=w$的条件下，S_n的条件概率分布列为

$$P(S_n=t|S_{n+1}=w)=\frac{P(S_n=t,S_{n+1}=w)}{P(S_{n+1}=w)}$$
$$=\frac{\binom{t-1}{n-1}p^{n+1}q^{w-(n+1)}}{\binom{w-1}{n}p^{n+1}q^{w-(n+1)}}$$
$$=\frac{\binom{t-1}{n-1}}{\binom{w-1}{n}},\quad t=n,\ldots,w-1.$$

这一条件分布不依赖于p.

II.Continuous case: $P(\xi = x) = 0$. Given $\xi = x$, the conditional distribution function of η can be understood as

$$P(\eta \leq y | \xi = x)$$

II.Continuous case: $P(\xi = x) = 0$. Given $\xi = x$, the conditional distribution function of η can be understood as

$$
\begin{aligned}
& P(\eta \le y | \xi = x) \\
= & \lim_{\Delta x \to 0} P(\eta \le y | x < \xi \le x + \Delta x) \\
= & \lim_{\Delta x \to 0} \frac{P(x < \xi \le x + \Delta x, \eta \le y)}{P(x < \xi \le x + \Delta x)} \\
= & \lim_{\Delta x \to 0} \frac{F(x + \Delta x, y) - F(x, y)}{F_\xi(x + \Delta x) - F_\xi(x)}.
\end{aligned}
$$

$$P(\eta \le y | \xi = x) = \frac{\partial F / \partial x}{F_\xi'(x)}$$

$$= \frac{\int_{-\infty}^{y} p(x,v)dv}{p_\xi(x)} = \int_{-\infty}^{y} \frac{p(x,v)}{p_\xi(x)} dv.$$

When $p_\xi(x) > 0$, conditioning on $\xi = x$, the density of η is

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}.$$

Definition

设随机向量(ξ,η)有联合密度函数$p(x,y)$, ξ有边际密度函数$p_\xi(x)=\int_{-\infty}^{\infty}p(x,y)dy$. 若在$x$ 处, $p_\xi(x)>0$, 则称

$$P(\eta\le y|\xi=x)=\int_{-\infty}^{y}\frac{p(x,v)}{p_\xi(x)}dv,\quad y\in\boldsymbol{R}$$

为在$\xi=x$的条件下, η的条件分布函数, 简称为条件分布, 记作$F_{\eta|\xi}(y|x)$. 称

$$p_{\eta|\xi}(y|x)=\frac{p(x,y)}{p_\xi(x)},\quad y\in\boldsymbol{R} \tag{1}$$

为在$\xi=x$的条件下, η的条件密度函数, 简称为条件密度.

若$p_\xi(x) = \int_{-\infty}^{\infty} p(x,y)dy = 0$，则对所有的$y$，$p(x,y) = 0$，(1)式右边是$\frac{0}{0}$型不定式，通常定义$p_{\eta|\xi}(y|x)$的值为0.

若$p_\xi(x) = \int_{-\infty}^{\infty} p(x,y)dy = 0$, 则对所有的$y$, $p(x,y) = 0$, (1)式右边是$\frac{0}{0}$型不定式, 通常定义$p_{\eta|\xi}(y|x)$的值为0.

同理, 若$p_\eta(y) > 0$, 则在$\eta = y$ 的条件下, ξ 的密度函数为

$$p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

Example

Suppose that $(\xi, \eta) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r)$, calculate the conditional density $p_{\eta|\xi}(y|x)$.

Example

Suppose that $(\xi,\eta)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,r)$, calculate the conditional density $p_{\eta|\xi}(y|x)$.

解 (ξ,η)的联合密度为

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

Example

Suppose that $(\xi, \eta) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r)$, calculate the conditional density $p_{\eta|\xi}(y|x)$.

解 (ξ, η)的联合密度为

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

下面我们推导在$\xi = x$的条件下，η的条件密度. 为此，我们不断把不含y的因子提出来，用常数C_i表示. 最后的常数通过$\int_{-\infty}^{\infty} p_{\eta|\xi}(y|x)dy = 1$求得.

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{\int_{-\infty}^{\infty} p(x,y)dy} = C_1 p(x,y)$$

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{\int_{-\infty}^{\infty} p(x,y)dy} = C_1 p(x,y)$$

$$= \quad C_2 \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(y-\mu_2)^2}{\sigma_2^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\}$$

$$
\begin{aligned}
p_{\eta|\xi}(y|x) &= \frac{p(x,y)}{\int_{-\infty}^{\infty} p(x,y)dy} = C_1 p(x,y) \\
&= C_2 \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(y-\mu_2)^2}{\sigma_2^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \\
&= C_3 \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{y-\mu_2}{\sigma_2} - r\frac{x-\mu_1}{\sigma_1}\right)^2\right\}
\end{aligned}
$$

$$
\begin{aligned}
p_{\eta|\xi}(y|x) &= \frac{p(x,y)}{\int_{-\infty}^{\infty} p(x,y)dy} = C_1 p(x,y) \\
&= C_2 \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(y-\mu_2)^2}{\sigma_2^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \\
&= C_3 \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{y-\mu_2}{\sigma_2} - r\frac{x-\mu_1}{\sigma_1}\right)^2\right\} \\
&= C_3 \exp\left\{-\frac{\left[y-\mu_2 - r\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2}{2\sigma_2^2(1-r^2)}\right\}.
\end{aligned}
$$

上述过程可以简写为

$$
\begin{aligned}
p_{\eta|\xi}(y|x) &\propto_y p(x,y) \propto_y \dots \\
&\propto_y \exp\left\{-\frac{\left[y-\mu_2-r\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2}{2\sigma_2^2(1-r^2)}\right\}.
\end{aligned}
$$

上述过程可以简写为

$$
\begin{aligned}
p_{\eta|\xi}(y|x) &\propto_y p(x,y) \propto_y \dots \\
&\propto_y \exp\left\{-\frac{\left[y-\mu_2-r\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2}{2\sigma_2^2(1-r^2)}\right\}.
\end{aligned}
$$

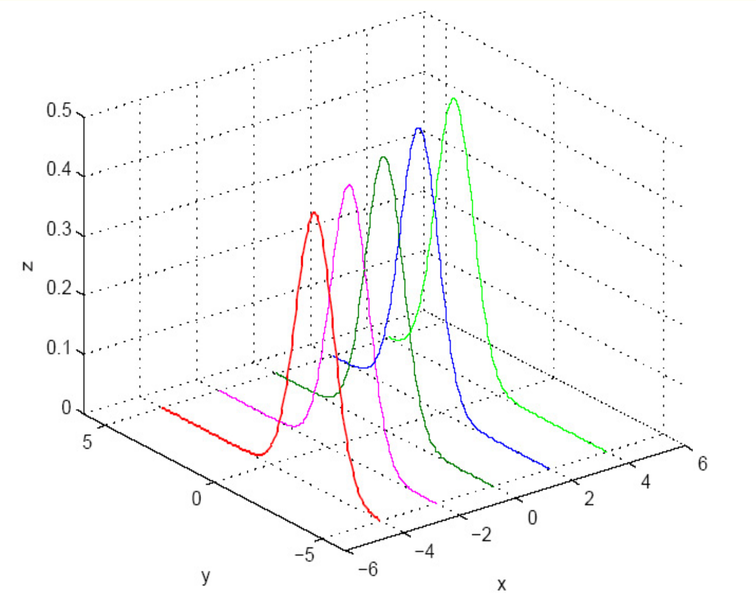
回顾正态分布的密度函数知$p_{\eta|\xi}(y|x)$为正态密度函数

$$
p_{\eta|\xi}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{\left[y-\mu_2-r\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2}{2\sigma_2^2(1-r^2)}\right\}. \tag{2}
$$

即在$\xi = x$的条件下，二维正态分布的条件分布是正态分布$N\left(\mu_2 + r\frac{\sigma_2}{\sigma_1}(x-\mu_1), (1-r^2)\sigma_2^2\right)$，记作

$$\eta\big|_{\xi=x} \sim N\Big(\underline{\mu_2 + \frac{r\sigma_2}{\sigma_1}(x-\mu_1)}, \underline{(1-r^2)\sigma_2^2}\Big),$$

其中第一个参数$m = \mu_2 + r\frac{\sigma_2}{\sigma_1}(x-\mu_1)$是$x$的线性函数，第二个参数与$x$无关.


0.5
0.4
0.3
0.2
0.1
0
z
5
0
−5
y
−6
−4
−2
0
2
4
6
x

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

If ξ and η are independent, then

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

If ξ and η are independent, then

$$p_{\eta|\xi}(y|x) = p_\eta(y), \quad p_{\xi|\eta}(x|y) = p_\xi(x).$$

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

$$p(x,y) = p_{\eta|\xi}(y|x)p_\xi(x); \quad p(x,y) = p_{\xi|\eta}(x|y)p_\eta(y).$$

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

$$p(x,y) = p_{\eta|\xi}(y|x)p_\xi(x); \quad p(x,y) = p_{\xi|\eta}(x|y)p_\eta(y).$$

$$\begin{aligned} p_{\xi|\eta}(x|y) &= \frac{p(x,y)}{p_\eta(y)} = \frac{p(x,y)}{\int p(u,y)du} \\ &= \frac{p_{\eta|\xi}(y|x)p_\xi(x)}{\int p_{\eta|\xi}(y|u)p_\xi(u)du}. \end{aligned}$$

$$p_{\eta|\xi}(y|x) = \frac{p(x,y)}{p_\xi(x)}; \quad p_{\xi|\eta}(x|y) = \frac{p(x,y)}{p_\eta(y)}.$$

$$p(x,y) = p_{\eta|\xi}(y|x)p_\xi(x); \quad p(x,y) = p_{\xi|\eta}(x|y)p_\eta(y).$$

$$\begin{aligned} p_{\xi|\eta}(x|y) &= \frac{p(x,y)}{p_\eta(y)} = \frac{p(x,y)}{\int p(u,y)du} \\ &= \frac{p_{\eta|\xi}(y|x)p_\xi(x)}{\int p_{\eta|\xi}(y|u)p_\xi(u)du}. \end{aligned}$$

Bayesian formula

III. The general case:

In general, suppose that the joint distribution function of (ξ, η) is $F(x, y)$. If

$$
\begin{aligned}
&\lim_{\Delta y \to 0} P(\xi \le x | \eta \in (y, y + \Delta y]) \\
= &\lim_{\Delta y \to 0} \frac{P(\xi \le x, \eta \in (y, y + \Delta y])}{P(\eta \in (y, y + \Delta y])} \\
= &\lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x, y)}{F_\eta(y + \Delta y) - F_\eta(y)}
\end{aligned}
$$

exists for any x, we call the limit function $F_{\xi|\eta}(x|y)$ be the conditional distribution function of ξ for given $\eta = y$.

If there exists $\{x_i\}$ such that

$$F_{\xi|\eta}(x|y) = \sum_{i:x_i \le x} p_{\xi|\eta}(x_i|y), \quad x \in \boldsymbol{R},$$

the we call $p_{\xi|\eta}(x_i|y)$, $i = 1, 2, \dots,$ the conditional mass function (条件分布列). If $F_{\xi|\eta}(x|y)$ can represented as the form

$$F_{\xi|\eta}(x|y) = \int_{-\infty}^{x} p_{\xi|\eta}(v|y)dv, \quad x \in \boldsymbol{R},$$

then we call $p_{\xi|\eta}(x|y)$ the conditional density function.

Example

设Λ服从伽玛分布$\Gamma(b,a)$, 在条件$\Lambda=\lambda$下, X服从参数为λ 的泊松分布. 求在$X=x$的条件下Λ 的分布.

Example

设Λ服从伽玛分布$\Gamma(b,a)$，在条件$\Lambda=\lambda$下，X服从参数为λ 的泊松分布. 求在$X=x$的条件下Λ 的分布.

解 Λ为连续型随机变量，X为离散型随机变量. 对$x=0,1,\ldots,$ 有

$$P(X=x|\Lambda=\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}.$$

Example

设Λ服从伽玛分布$\Gamma(b,a)$，在条件$\Lambda=\lambda$下，X服从参数为λ 的泊松分布. 求在$X=x$的条件下Λ 的分布.

解 Λ为连续型随机变量，X为离散型随机变量. 对$x=0,1,\dots,$有

$$P(X=x|\Lambda=\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}. \quad + o(\Delta\lambda)$$

这意味着

$$P(X=x|\Lambda=\lambda)=\lim_{\Delta\lambda\to 0}\frac{P\big(X=x,\Lambda\in(\lambda,\lambda+\Delta\lambda)\big)}{P\big(\Lambda\in(\lambda,\lambda+\Delta\lambda)\big)}$$

即

$$
\begin{aligned}
& P\big(X = x, \Lambda \in (\lambda, \lambda + \Delta\lambda)\big) \\
=& P(X = x | \Lambda = \lambda) P\big(\Lambda \in (\lambda, \lambda + \Delta\lambda)\big) + o(\Delta\lambda) \\
=& P(X = x | \Lambda = \lambda) p_\Lambda(\lambda) \Delta\lambda + o(\Delta\lambda).
\end{aligned}
$$

即

$$
\begin{aligned}
& P\big(X = x, \Lambda \in (\lambda, \lambda + \Delta\lambda)\big) \\
=& P(X = x | \Lambda = \lambda) P\big(\Lambda \in (\lambda, \lambda + \Delta\lambda)\big) + o(\Delta\lambda) \\
=& P(X = x | \Lambda = \lambda) p_\Lambda(\lambda) \Delta\lambda + o(\Delta\lambda).
\end{aligned}
$$

所以

$$
P(X = x, \Lambda \le y) = \int_{-\infty}^{y} P(X = x | \Lambda = \lambda) p_\Lambda(\lambda) d\lambda.
$$

即

$$
\begin{aligned}
& P\big(X=x, \Lambda \in (\lambda, \lambda+\Delta\lambda)\big) \\
=& P(X=x|\Lambda=\lambda) P\big(\Lambda \in (\lambda, \lambda+\Delta\lambda)\big) + o(\Delta\lambda) \\
=& P(X=x|\Lambda=\lambda) p_\Lambda(\lambda)\Delta\lambda + o(\Delta\lambda).
\end{aligned}
$$

所以

$$
P(X=x, \Lambda \le y) = \int_{-\infty}^{y} P(X=x|\Lambda=\lambda) p_\Lambda(\lambda) d\lambda.
$$

从而

$$
\begin{aligned}
P(\Lambda \le y|X=x) =& \frac{P(X=x, \Lambda \le y)}{P(X=x)} \\
=& \int_{-\infty}^{y} \frac{P(X=x|\Lambda=\lambda) p_\Lambda(\lambda)}{P(X=x)} d\lambda.
\end{aligned}
$$

因此在$X = x$的条件下，Λ的密度函数为

$$
\begin{aligned}
p_{\Lambda|\xi}(\lambda|x) &= \frac{P(X = x|\Lambda = \lambda)p_\Lambda(\lambda)}{P(X = x)} \\
&= \frac{P(X = x|\Lambda = \lambda)p_\Lambda(\lambda)}{\int_{-\infty}^{\infty} P(X = x|\Lambda = \lambda)p_\Lambda(\lambda)d\lambda} \\
&\propto_\lambda \lambda^x e^{-\lambda} \lambda^{b-1} e^{-\lambda a} = \lambda^{x+b-1} e^{-(a+1)\lambda}, \quad \lambda > 0.
\end{aligned}
$$

因此在$X=x$的条件下，Λ的密度函数为

$$p_{\Lambda|\xi}(\lambda|x)=\frac{P(X=x|\Lambda=\lambda)p_\Lambda(\lambda)}{P(X=x)}$$
$$=\frac{P(X=x|\Lambda=\lambda)p_\Lambda(\lambda)}{\int_{-\infty}^{\infty}P(X=x|\Lambda=\lambda)p_\Lambda(\lambda)d\lambda}$$
$$\propto_\lambda \lambda^x e^{-\lambda}\lambda^{b-1}e^{-\lambda a}=\lambda^{x+b-1}e^{-(a+1)\lambda},\quad \lambda>0.$$

将上式右边添加正则化常数因子使得其积分为1，得

$$p_{\Lambda|\xi}(\lambda|x)=\frac{(a+1)^{x+b}}{\Gamma(x+b)}\lambda^{x+b-1}e^{-(a+1)\lambda},\quad \lambda>0.$$

即在$X=x$的条件下，Λ服从伽玛分布$\Gamma(x+b,a+1)$.

$$
\begin{aligned}
p_{\Lambda|X}(\lambda|x) &= \frac{P(X=x|\Lambda=\lambda)p_\Lambda(\lambda)}{P(X=x)} \\
&= \frac{P(X=x|\Lambda=\lambda)p_\Lambda(\lambda)}{\int_{-\infty}^{\infty} P(X=x|\Lambda=\lambda)p_\Lambda(\lambda)d\lambda}
\end{aligned}
$$

Bayes 公式.

IV. Multi-dimensional case:

Suppose that the joint distribution function of random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ is $F(\boldsymbol{x}, \boldsymbol{y})$. If

$P(\xi \le x \mid \eta = y)$

$$\lim_{\Delta \boldsymbol{y}\to 0} P(\boldsymbol{\xi} \le \boldsymbol{x} | \boldsymbol{\eta} \in (\boldsymbol{y}, \boldsymbol{y}+\Delta \boldsymbol{y}])$$

$$= \lim_{\Delta \boldsymbol{y}\to 0} \frac{P(\boldsymbol{\xi} \le \boldsymbol{x}, \boldsymbol{\eta} \in (\boldsymbol{y}, \boldsymbol{y}+\Delta \boldsymbol{y}])}{P(\boldsymbol{\eta} \in (\boldsymbol{y}, \boldsymbol{y}+\Delta \boldsymbol{y}]}$$

$$= \lim_{\Delta \boldsymbol{y}\to 0} \frac{F(\boldsymbol{x}, \boldsymbol{y}+\Delta \boldsymbol{y}) - F(\boldsymbol{x}, \boldsymbol{y})}{F_{\eta}(\boldsymbol{y}+\Delta \boldsymbol{y}) - F_{\eta}(\boldsymbol{y})}$$

exists for any $\boldsymbol{x}$, we call the limit function $F_{\xi|\eta}(\boldsymbol{x}|\boldsymbol{y})$ be the conditional distribution function of $\boldsymbol{\xi}$ for given $\boldsymbol{\eta} = \boldsymbol{y}$.

When $(\boldsymbol{\xi}, \boldsymbol{\eta})$ is a continuous random vector with probability density function $p(\boldsymbol{x}, \boldsymbol{y})$, the conditional probability density function of $\boldsymbol{\xi}$ for given $\boldsymbol{\eta} = \boldsymbol{y}$ is

$$p_{\xi|\eta}(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p_\eta(\boldsymbol{y})} = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{\int p(\boldsymbol{u}, \boldsymbol{y}) d\boldsymbol{u}},$$

$$\text{if } \; p_\eta(\boldsymbol{y}) > 0.$$

When $(\boldsymbol{\xi}, \boldsymbol{\eta})$ is a discrete random vector with probability mass function $P(\boldsymbol{\xi} = \boldsymbol{x}_i, \boldsymbol{\eta} = \boldsymbol{y}_j) = p(\boldsymbol{x}_i, \boldsymbol{y}_j)$, the conditional probability mass function of $\boldsymbol{\xi}$ for given $\boldsymbol{\eta} = \boldsymbol{y}_j$ is

$$p_{\boldsymbol{\xi}|\boldsymbol{\eta}}(\boldsymbol{x}_i|\boldsymbol{y}_j) = \frac{p(\boldsymbol{x}_i, \boldsymbol{y}_j)}{p_{\boldsymbol{\eta}}(\boldsymbol{y}_j)} = \frac{P(\boldsymbol{\xi} = \boldsymbol{x}_i, \boldsymbol{\eta} = \boldsymbol{y}_j)}{P(\boldsymbol{\eta} = \boldsymbol{y}_j)},$$

$$\text{if } P(\boldsymbol{\eta} = \boldsymbol{y}_j) > 0.$$