Note 05 - Jan 24

3. Conditional distribution and conditional expectation

3.1 Conditional distribution

3.1.1 Discrete case

Definition Let X and Y be discrete r.v's. The conditional distribution of X given Y is given by:

$$P(X=x|Y=y)=\frac{(P(X=x,Y=u))}{P(Y=y)}$$

$$P(X=x|Y=y): f_{X|Y}=y(x), f_{X|Y}(x|y) \leftarrow \text{conditional probability mass function)}$$

Conditional pmf is a legitimate pmf: given any y, $f_{X|Y=y}(x) \geq 0, \forall x$

$$\sum_x f_{X|Y=y}(x)=1$$

Note that given $Y=y$, as x changes, the value of the function $f_{X|Y=y}(x)$ is proportional to the joint probability.

$$f_{X|Y=y}(x) \propto P(X=x,Y=y)$$

This is useful for solving problems where the denominator $P(Y=y)$ is hard to find.

3.1.1.1 Example

$X_1 \sim Poi(\lambda_1), X_2 \sim Poi(\lambda_2)$. $X_1 \perp\!\!\!\perp X_2, Y=X_1+X_2$

Q: $P(X_1=k|Y=n)$?

Note $P(X_1=k|Y=u)=f_{X_1|Y=n}(k)$

A: $P(X_1=k|Y=n)$ can only be non-zero for $k=0,\cdots,n$ in this case,

$$
\begin{aligned}
P(X_1 = k|Y = n) &= \frac{P(X_1 = k, Y = n)}{P(Y = n)} \\
&\propto P(X_1 = k, Y = n) \\
&= P(X_1 = k, X_2 = n - k) \\
&= e^{-\lambda_1} \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\
&\propto (\frac{\lambda_1}{\lambda_2})^k / k!(n-k)!
\end{aligned}
$$

we can get $P(X = k|Y = n)$ by normalizing the above expression.

$$
P(X_1 = k, Y = n) = \frac{(\frac{\lambda_1}{\lambda_2})^k / k!(n-k)!}{\sum_{k=0}^{n} (\frac{\lambda_1}{\lambda_2})^k / k!(n-k)!}
$$

but then we will need to fine $\sum_{k=0}^{n} (\frac{\lambda_1}{\lambda_2})^k / k!(n-k)!$

An easier way is to compare $\sum_{k=0}^{n} (\frac{\lambda_1}{\lambda_2})^k / k!(n-k)!$ with the known results for common distribution. In particular, if $X \sim Bin(n, p)$

$$
\begin{aligned}
P(X = k) &= (\overset{n}{k}) p^k (1-p)^{n-k} \\
&\propto (\frac{p}{1-p})^k / k!(n-k)!
\end{aligned}
$$

$\Rightarrow P(X_1 = k|Y = n)$ follows a binomial distributions with parameters n and p given by $\frac{p}{1-p} = \frac{\lambda_1}{\lambda_2} \Rightarrow p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$

Thus, given $Y = X_1 + X_2 = n$, the conditional distribution of X_1 is binomial with parameter n and $\frac{\lambda_1}{\lambda_1 + \lambda_2}$

3.1.2 Continuous case

Definition: Let X and Y be continuous r.v's. The conditional distribution of X given Y is given by

$$
f_{X|Y}(x|y) = f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}
$$

A conditional pdf is a legitimate pdf

$$
\begin{aligned}
f_{X|Y}(x|y) \geq 0 \qquad & x, y \in \mathbb{R} \\
\int_{-\infty}^{\infty} f_{X|Y}(x|y) dx = 1, \quad & y \in \mathbb{R}
\end{aligned}
$$

3.1.2.1 Example

Suppose $X \sim Exp(\lambda), Y|X = x \sim Exp(x) = f_{Y|X}(y|x) = xe^{-xy}, y = e \leftarrow$ conditional distribution of Y given $X = x$

Q: Find the condition pdf $f_{X|Y}(x|y)$

A:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &\propto f(x,y) \\ &= f_{Y|X}(y|x) \cdot f_X(x) \\ &= xe^{xy}\lambda e^{-\lambda xa} \\ &\propto xe^{-x(y+\lambda)}, \qquad x > 0, y > 0 \end{aligned}$$

Normalization (make the total probability 1)

$$f_{X|Y}(x|y) = \frac{xe^{-x(y+\lambda)}}{\int_0^\infty xe^{-x(y+\lambda)}dx}$$

$$\int_0^\infty xe^{-x(y+\lambda)}dx = \frac{1}{\lambda + t}^2 \leftarrow \text{integration by parts}$$

Thus, $f_{X|Y}(x|y) = (\lambda + y)^2 xe^{-x(y+\lambda)}$, $x > 0$.

This is a gamma distribution with parameters γ and $\lambda + y$

3.1.2.1. Example 2

Find the distribution of $z = XY$.

Attention: the following method is wrong:

$$f_Z(z) = \int_0^\infty f_{Y|X}(\frac{z}{x}|x) \cdot f_X(x)dx$$

If we want to directly work with pdf's, we will need to use the change of variable formula for multi-variables. The right formula have turns out to be

$$\begin{aligned} f_Z(z) &= \int_0^\infty f_{X,Z}(x,z)dx = \int_0^\infty f_{Z|X}(z|x)f_X(x)dx \\ &= \int_0^\infty f(x, \frac{z}{x}) \cdot \frac{1}{x}dx \\ &= f_{Y|X}(\frac{z}{x}|x)f_X(x) \cdot \frac{1}{x}dx \end{aligned}$$

As an **easier way** is to use cdf, which gives probability rather than density:

$$
\begin{aligned}
P(Z=z) &= P(XY \le z) \\
&= \int_0^\infty P(XY \le z | X = x) f_X(x) dx \qquad \text{(law of total probability)} \\
&= \int_0^\infty P(Y \le \frac{z}{x} | X = x) \cdot f_X(x) dx
\end{aligned}
$$

$$Y|X = x \sim Exp(x)$$

$$
\begin{aligned}
&= \int_0^\infty (1 - e^{-x \cdot \frac{z}{x}}) \cdot \lambda e^{-\lambda x} dx \\
&= 1 - e^{-z} \int_0^\infty \lambda e^{-\lambda x} dx
\end{aligned}
$$

$$\Rightarrow Z \sim Exp(1)$$

Notation $X, Y | \{Z = k\} \overset{iid}{\sim} \cdots$ means that given $Z = k$, X and Y are *conditionally independent*, and they follow certain distribution.

(the conditional joint cdf/pmf/pdf equals the predict of the conditional cdf's/pmf's/pdf's)

3.2 Conditional expectation

We have seen that conditional pmf/pdf are legitimate pmf/pdf. Correspondingly, a conditional distribution is nothing else but a probability distributions. It is simply a (potentially) different distribution, since it takes more information into consideration.

As a result, we can define everything which are previously defined for unconditional distributions also for conditional distributions.

In particular, it is natural to define the conditional expectation.

Definition. The conditional expectation of $g(X)$ given $Y = y$ is defined as

$$\mathbb{E}(g(X)|Y = y) = \begin{cases} \sum_{i_1}^\infty g(x_i) P(X = x_u | Y = y) & \text{if } X|Y = y \text{ is discrete} \\ \int_{-\infty}^\infty g(x) f_{X|Y}(x|y) dx & \text{if } X|X = y \text{ is continuous} \end{cases}$$

Fix y, the conditional expectation is nothing but the expectation taken under the conditional distribution.