Statistical Risk Analysis
Chapter 5: Conditional Distributions with Applications

Shaobo Jin

Department of Mathematics

Dependence

We often assume the measurements are independent. However, if we have several measurements on the same individual, the measurements are often dependent.

- Suppose that we have two random variables X and Y. The distribution function is

$$F_{X,Y}(x,y) = \mathrm{P}(X \leq x, Y \leq y).$$

  It is often called the joint distribution.

- The marginal distributions are

$$\begin{aligned} F_X(x) &= \mathrm{P}(X \leq x) = F_{X,Y}(x,\infty), \\ F_Y(y) &= \mathrm{P}(Y \leq y) = F_{X,Y}(\infty,y). \end{aligned}$$

Probability Mass Function

If X and Y take a finite or countable number of values (e.g., 0, 1, 2, ...), then the distribution

$$F_{X,Y}(x, y) = \sum_{j \leq x} \sum_{k \leq y} p_{jk},$$

where $p_{jk} = P(X = j, Y = k)$ is the joint probability mass function (pmf).

The marginal probability mass function can be computed by

$$P(X = j) = \sum_{k=0}^{\infty} \mathrm{P}(X = j, Y = k),$$

$$P(Y = k) = \sum_{j=0}^{\infty} \mathrm{P}(X = j, Y = k).$$

Example: A Multinomial Distribution

Consider a multinomial distribution of three outcomes. Let X be the number of the first outcome and Y be the number of the second outcome after n trials. Its pmf is

$$\mathrm{P}\left(X=j,Y=k\right) = \frac{n!}{j!k!\left(n-j-k\right)!}p_A^j p_B^k\left(1-p_A-p_B\right)^{n-j-k}$$

where $0\leq p_A\leq 1$ and $0\leq p_B\leq 1$ are the probabilities of the first and second outcome respectively.
The marginal is

$$\begin{aligned}\mathrm{P}\left(X=j\right) &= \frac{n!}{j!\left(n-j\right)!}p_A^j\left(1-p_A\right)^{n-j},\\ \mathrm{P}\left(Y=k\right) &= \frac{n!}{k!\left(n-k\right)!}p_B^k\left(1-p_B\right)^{n-k}.\end{aligned}$$

That is, $X\in\mathrm{Bin}\left(n,p_A\right)$ and $Y\in\mathrm{Bin}\left(n,p_B\right)$.

Probability Density Function

If $F_{X,Y}(x,y)$ is differentiable with respect to x and y, the derivative

$$f(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$$

is called the joint probability density function (pdf). Then

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dtds, \\ P(X \in A, Y \in B) &= \int_{x \in A}\int_{y \in B} f(s,t)\,dtds. \end{aligned}$$

The marginal densities of X and Y are obtained by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,t)\,dt \qquad f_Y(y) = \int_{-\infty}^{\infty} f(s,y)\,ds.$$

Example: Bivariate Normal Distribution

For $\rho \in (-1, 1)$, the joint density of a bivariate normal distribution is

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

where

$$u = \frac{(x - m_X)^2}{\sigma_X^2} - \frac{2\rho(x - m_X)(y - m_y)}{\sigma_X\sigma_Y} + \frac{(y - m_y)^2}{\sigma_Y^2}.$$

We denote it by $(X, Y) \in N\left(m_X, m_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$.

The marginal is $X \in N\left(m_X, \sigma_X^2\right)$ and $Y \in N\left(m_Y, \sigma_Y^2\right)$.

Expectation

Let $h(X,Y)$ be a function of X and Y. Then,

$$E[h(X,Y)] = \begin{cases} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} h(j,k)\, p_{jk}, & (X,Y) \text{ is discrete,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)\, f(x,y)\, dx dy, & (X,Y) \text{ is continuous.} \end{cases}$$

In the special case where $h(X,Y) = aX + bY$ for constants a and b, we have

$$E[aX + bY] = aE[X] + bE[Y].$$

Covariance

The covariance between X and Y is defined by

$$\operatorname{Cov}[X,Y] = E[XY] - E[X]E[Y].$$

1. Covariance is symmetric:

$$\operatorname{Cov}[Y,X] = E[YX] - E[Y]E[X] = \operatorname{Cov}[X,Y].$$

2. The covariance between X and X is the variance:

$$\operatorname{Cov}[X,X] = E\left[X^2\right] - E[X]E[X] = V(X).$$

3. The covariance matrix of (X,Y) is

$$\begin{bmatrix} V[X] & \operatorname{Cov}[X,Y] \\ \operatorname{Cov}[X,Y] & V[Y] \end{bmatrix}.$$

Covariance In Variance

Variance of a linear combination is

$$V\left[aX+bY+c\right] = a^2V\left[X\right]+2ab\mathrm{Cov}\left[X,Y\right]+b^2V\left[Y\right].$$

In general,

$$\begin{aligned} V\left[\sum_{i=1}^{n}a_iX_i+c\right] &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}\left[X_i,X_j\right] \\ &= \sum_{i=1}^{n}a_i^2V\left[X_i\right]+\sum_{i=1}^{n}\sum_{j\neq i}\mathrm{Cov}\left[X_i,X_j\right] \\ &= \sum_{i=1}^{n}a_i^2V\left[X_i\right]+2\sum_{i=2}^{n}\sum_{j<i}\mathrm{Cov}\left[X_i,X_j\right]. \end{aligned}$$

Pearson Correlation

The correlation between X and Y is

$$\rho_{XY} = \frac{\operatorname{Cov}[X, Y]}{D[X]\, D[Y]} \in [-1, 1],$$

where $D[X] = \sqrt{V[X]}$ and $D[Y] = \sqrt{V[Y]}$.

An estimate is

$$\rho^*_{XY} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}.$$

Examples

1. Consider

$$\mathrm{P}(X=j, Y=k) = \frac{n!}{j!k!(n-j-k)!} p_A^j p_B^k (1-p_A-p_B)^{n-j-k}.$$

The covariance between X and Y is

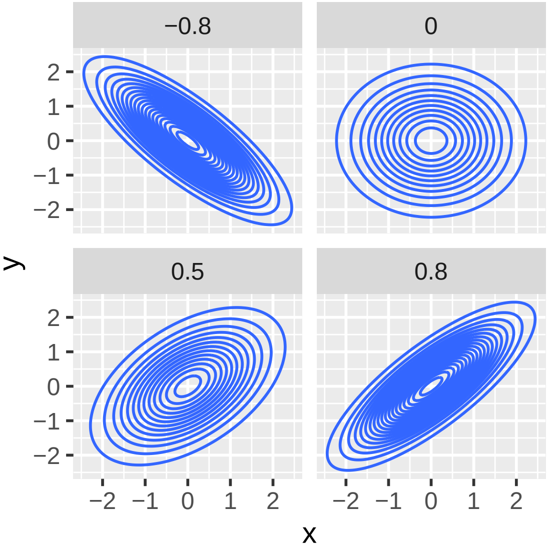
$$\mathrm{Cov}[X, Y] = -np_A p_B.$$

2. Consider $(X, Y) \in N\left(m_X, m_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$.

$$\mathrm{Cov}[X, Y] = \rho\sigma_X\sigma_Y,$$

and ρ is the correlation between X and Y.

Contour of Bivariate Normal Density



Application of Bivariate Normal

Let ℓ be the log-likelihood function.

1. Suppose that we want to estimate one parameter θ. The distribution of $\left[\ddot{\ell}\left(\theta^*\right)\right]^{-1/2}\left(\Theta^* - \theta\right)$ can often be approximated by $N\left(0,1\right)$.
2. Suppose that we want to estimate two parameters θ_1 and θ_2. The distribution of can often be approximated by

$$\left[\ddot{\ell}\left(\theta_1^*, \theta_2^*\right)\right]^{-1/2}\left(\begin{bmatrix}\Theta_1^* \\ \Theta_2^*\end{bmatrix} - \begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix}\right) \in AsN\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}1 & 0 \\ 0 & 1\end{bmatrix}\right).$$

The covariance matrix of Θ_1^* and Θ_2^* can be approximated by

$$-\begin{bmatrix}\frac{\partial^2 \ell\left(\theta_1^*, \theta_2^*\right)}{\partial \theta_1^2} & \frac{\partial^2 \ell\left(\theta_1^*, \theta_2^*\right)}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \ell\left(\theta_1^*, \theta_2^*\right)}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ell\left(\theta_1^*, \theta_2^*\right)}{\partial \theta_2^2}\end{bmatrix}^{-1}.$$

Example: MLE of Multinomial Distribution

Suppose that we have data from a multinomial distribution with pmf

$$\mathrm{P}\left(X=j, Y=k\right) = \frac{n!}{j!k!\left(n-j-k\right)!}p_A^j p_B^k \left(1-p_A-p_B\right)^{n-j-k}.$$

Approximate the distribution of p_A^* and p_B^*.

Conditioning

The conditional probability is

$$\mathrm{P}(B \mid A) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(A)}.$$

1. The conditional probability mass function is

$$\mathrm{P}(X = j \mid Y = k) = \frac{\mathrm{P}(X = j, Y = k)}{\mathrm{P}(Y = k)}.$$

   We have $\sum_{j=-\infty}^{\infty} \mathrm{P}(X = j \mid Y = k) = 1$.

2. The conditional probability density function is

$$f(x \mid y) = \frac{f(x, y)}{f(y)}, \quad \text{if } f(y) > 0 \text{ and zero otherwise.}$$

   We have $F(x \mid y) = \int_{-\infty}^{x} f(t \mid y)\, dt$.

Law of Total Probability

Theorem (Theorem 5.2 and 5.3, Law of Total Probability)

Consider an event B. Then,

$$P(B) = \begin{cases} \sum_{i=0}^{\infty} P(B \mid Y = i) P(Y = i) & Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} P(B \mid Y = y) f_Y(y)\, dy & Y \text{ is continuous.} \end{cases}.$$

If X and Y have joint density $f(x, y)$ and B is a statement about X, then

$$P(B \mid Y = y) = \int_{x \in B} f(x \mid y)\, dx.$$

Example: Number of Cracks

1. Suppose that number of cracks of a tanker follows a Poisson distribution with mean m. A devices can detect a crack with probability 0.999. Let $B = \{\text{All creacks have been detected}\}$. Find $\mathrm{P}(B)$.
2. Suppose that the strength of concrete M follows a normal distribution $N\left(\mu, \sigma^2\right)$. Suppose that, given the strength m, the number of cracks of a product made by such concrete follows a Poisson distribution with mean $m(1-m)$. Find the probability that the number of cracks is 2.
3. Suppose that we have two fuses. The times that the fuses last are X and Y respectively. We assume that they follow independent exponential distribution with mean θ. Find the probability that $X \leq Y$.

Independence

Suppose that two random variables X and Y are independent. Then,

1. $\text{Cov}\left[X, Y\right] = 0$ and $E\left[XY\right] = E\left[X\right]E\left[Y\right]$.
2. If X and Y are discrete random variables, then

$$
\begin{aligned}
p_{jk} &= \mathrm{P}\left(X = j\right)\mathrm{P}\left(Y = k\right), \\
\mathrm{P}\left(X = j \mid Y = k\right) &= \mathrm{P}\left(X = j\right).
\end{aligned}
$$

3. If X and Y are discrete continuous variables, then

$$
\begin{aligned}
f\left(x, y\right) &= f_X\left(x\right)f_Y\left(y\right), \\
f\left(x \mid Y = y\right) &= f\left(x\right).
\end{aligned}
$$

Bayes' Formula

- For events, we have seen the Bayes' formula

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\,\mathrm{P}(A)}{\mathrm{P}(B)}.$$

- The conditional cdf of Y given event B is $P(Y \leq y \mid B)$.
- The pdf or pmf of this conditional distribution is

$$f_{Y|B}(y) = \frac{\mathrm{P}(B \mid Y = y)\, f_Y(y)}{\mathrm{P}(B)},$$

where

$$\mathrm{P}(B) = \begin{cases} \sum_{y} \mathrm{P}(B \mid Y = y)\,\mathrm{P}(Y = y), & Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} \mathrm{P}(B \mid Y = y)\, f_Y(y)\, dy, & Y \text{ is continuous.} \end{cases}$$

Example: Bayes' Formula

1. Let X be the maximal weight that will be carried by the wire for a period of 1 year. Assume that

$$\mathrm{P}\left(X \leq x\right) = \exp\left\{-e^{-(x-b)/a}\right\}, \quad x \in \mathbb{R},$$

   where $a = 156$ and $b = 910$.

2. Let Y be the strength of the wire (capacity to carry a load). We assume that Y follows a Weibull distribution with the density

$$f\left(y\right) = \frac{c}{\alpha}\left(\frac{y}{\alpha}\right)^{c-1}\exp\left\{-\left(\frac{y}{\alpha}\right)^{c}\right\}, \quad y \geq 0,$$

   where $c = 5.79$ and $\alpha = 1080$.

3. We also assume that X and Y are independent.

Let $B = \{\text{Safe operation during 1 year}\}$. Find $\mathrm{P}\left(B\right)$.

Example: Continue

Suppose that the wire has been used for one year and event B is true. We can update our knowledge about the wire to

$$
\begin{aligned}
f_{Y|B}(y) &= \frac{P(B \mid Y = y) f_Y(y)}{P(B)} \\
&= \frac{\exp\left\{-e^{-(x-910)/156}\right\} \cdot \frac{5.79}{1080}\left(\frac{y}{1080}\right)^{4.79} \exp\left\{-\left(\frac{y}{1080}\right)^{5.79}\right\}}{0.533}
\end{aligned}
$$

Let $C = \{\text{Safe operation during the next year}\}$. Hence,

$$
P(C) = \int_0^\infty P(C \mid Y = y) f_{Y|B}(y)\, dy \approx 0.705.
$$

Conditional independence

Denote by X_1 and X_2 the maximal load during the first and second year, respectively. We assume have that X_1 and X_2 are independent. Let $B_1 = \{X_1 < Y\}$ and $B_2 = \{X_2 > Y\}$.

1. If B_1 and B_2 were independent, then

$$P\left(\text{the wire survives two years}\right) = P\left(B_1\right) P\left(B_2\right) = (0.533)^2 .$$

2. If B_1 and B_2 were not independent, we may only have

$$P\left(B_1 \cap B_2 \mid Y = y\right) = P\left(B_1 \mid Y = y\right) P\left(B_2 \mid Y = y\right)$$

Then,

$$\begin{aligned} P\left(\text{the wire survives two years}\right) &= P\left(B_2 \mid B_1\right) P\left(B_1\right) \\ &= 0.705 \cdot 0.533 \end{aligned}$$