Proability and Statistics

MA20205

Bibhas Adhikari

Autumn 2022-23, IIT Kharagpur

Lecture 12
October 11, 2022

Joint distributions

Joint expectation

Let X, Y be random variables. Then the joint expectation of the pair is defined as

$$E(XY) = \begin{cases} \sum_{y\in\Omega_Y} \sum_{x\in\Omega_X} xy\, f(x,y) \text{ if } X, Y \text{ are discrete} \\ \int_{y\in\Omega_Y} \int_{x\in\Omega_X} xy\, f(x,y)\, dx\, dy \text{ if } X, Y \text{ are continuous} \end{cases}$$

Joint distributions

Joint expectation

Let X, Y be random variables. Then the joint expectation of the pair is defined as

$$E(XY) = \begin{cases} \sum_{y\in\Omega_Y} \sum_{x\in\Omega_X} xy\, f(x,y) \text{ if } X, Y \text{ are discrete} \\ \\ \int_{y\in\Omega_Y} \int_{x\in\Omega_X} xy\, f(x,y)\, dx\, dy \text{ if } X, Y \text{ are continuous} \end{cases}$$

Question Why is the joint expectation defined as the product random variable instead of addition ($E(X+Y)$) or difference ($E(X-Y)$) or quotient ($E(X/Y)$)?

Joint distributions

Suppose X and Y are discrete random variables with range spaces $\Omega_X = \{x_1, \ldots, x_n\}$ and $\Omega_Y = \{y_1, \ldots, y_n\}$.

Joint distributions

Suppose X and Y are discrete random variables with range spaces $\Omega_X = \{x_1, \ldots, x_n\}$ and $\Omega_Y = \{y_1, \ldots, y_n\}$. Now define the vectors

$$\boldsymbol{x} = [x_1, \ldots, x_n]^T, \ \boldsymbol{y} = [y_1, \ldots, y_n]^T.$$

Joint distributions

Suppose X and Y are discrete random variables with range spaces $\Omega_X = \{x_1, \ldots, x_n\}$ and $\Omega_Y = \{y_1, \ldots, y_n\}$. Now define the vectors

$$\boldsymbol{x} = [x_1, \ldots, x_n]^T, \; \boldsymbol{y} = [y_1, \ldots, y_n]^T.$$

Then define the pmf matrix as

$$\boldsymbol{P} = \begin{bmatrix} f(x_1, y_1) & f(x_1, y_2) & \ldots & f(x_1, y_n) \\ f(x_2, y_1) & f(x_2, y_2) & \ldots & f(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_n, y_1) & f(x_n, y_2) & \ldots & f(x_n, y_n) \end{bmatrix}.$$

Joint distributions

Suppose X and Y are discrete random variables with range spaces $\Omega_X = \{x_1, \ldots, x_n\}$ and $\Omega_Y = \{y_1, \ldots, y_n\}$. Now define the vectors

$$\boldsymbol{x} = [x_1, \ldots, x_n]^T, \; \boldsymbol{y} = [y_1, \ldots, y_n]^T.$$

Then define the pmf matrix as

$$\boldsymbol{P} = \begin{bmatrix} f(x_1, y_1) & f(x_1, y_2) & \ldots & f(x_1, y_n) \\ f(x_2, y_1) & f(x_2, y_2) & \ldots & f(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(x_n, y_1) & f(x_n, y_2) & \ldots & f(x_n, y_n) \end{bmatrix}.$$

Then

$$E(XY) = \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{y},$$

the weighted inner (scalar) product of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Joint distributions

For example, if $\Omega_X = \{1, \ldots, n\} = \Omega_Y$ with

$$f(x, y) = \begin{cases} \frac{1}{n} \, x = y \\ 0 \, x \neq y \end{cases}$$

then

$$\boldsymbol{P} = \frac{1}{n}\boldsymbol{I}, \; E(XY) = \frac{1}{n}\boldsymbol{x}^T\boldsymbol{y}$$

Joint distributions

Recall that the cosine angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$\cos\theta = \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

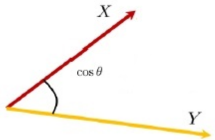
where $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$ and $\|\boldsymbol{y}\| = \sqrt{\sum_{i=1}^n y_i^2}$.

Joint distributions

Recall that the cosine angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}$$

where $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}$ and $\|\boldsymbol{y}\| = \sqrt{\sum_{i=1}^{n} y_i^2}$.
Geometry of expectation: geometry defined by the weighted inner product and weighted norm



where

$$\cos\theta = \frac{\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{y}}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} = \frac{E(XY)}{\sqrt{E(X^2)}\,\sqrt{E(Y^2)}}$$

Joint distribution

In the above,

$$\begin{aligned} E(X^2) &= \boldsymbol{x}^T \boldsymbol{P}_X \boldsymbol{x} = \|\boldsymbol{x}\|^2_{\boldsymbol{P}_X} \\ E(Y^2) &= \boldsymbol{y}^T \boldsymbol{P}_Y \boldsymbol{y} = \|\boldsymbol{y}\|^2_{\boldsymbol{P}_Y} \end{aligned}$$

where

$$\boldsymbol{P}_X = \begin{bmatrix} p(x_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p(x_n) \end{bmatrix}, \boldsymbol{P}_Y = \begin{bmatrix} p(y_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p(y_n) \end{bmatrix}$$

Joint distribution

In the above,

$$\begin{aligned} E(X^2) &= \boldsymbol{x}^T \boldsymbol{P}_X \boldsymbol{x} = \|\boldsymbol{x}\|^2_{\boldsymbol{P}_X} \\ E(Y^2) &= \boldsymbol{y}^T \boldsymbol{P}_Y \boldsymbol{y} = \|\boldsymbol{y}\|^2_{\boldsymbol{P}_Y} \end{aligned}$$

where

$$\boldsymbol{P}_X = \begin{bmatrix} p(x_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p(x_n) \end{bmatrix}, \boldsymbol{P}_Y = \begin{bmatrix} p(y_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p(y_n) \end{bmatrix}$$

Obviously,

$$-1 \leq \frac{E(XY)}{\sqrt{E(X^2)}\sqrt{E(Y^2)}} \leq 1$$

due to Cauchy-Schwarz inequality:

$$(E(XY))^2 \leq E(X^2)E(Y^2)$$

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Covariance

Let X, Y be random variables with means μ_X, μ_Y respectively. Then

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Covariance

Let X, Y be random variables with means μ_X, μ_Y respectively. Then

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Note: If $X = Y$ then $\text{Cov}(X, Y) = \text{Var}(X)$. Hence, covariance is a generalization of variance.

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Covariance

Let X, Y be random variables with means μ_X, μ_Y respectively. Then

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Note: If $X = Y$ then $\text{Cov}(X, Y) = \text{Var}(X)$. Hence, covariance is a generalization of variance.

Observations:

1. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Covariance

Let X, Y be random variables with means μ_X, μ_Y respectively. Then

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Note: If $X = Y$ then $\text{Cov}(X, Y) = \text{Var}(X)$. Hence, covariance is a generalization of variance.
Observations:

1. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
2. $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Covariance

Let X, Y be random variables with means μ_X, μ_Y respectively. Then

$$\mathrm{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Note: If $X = Y$ then $\mathrm{Cov}(X, Y) = \mathrm{Var}(X)$. Hence, covariance is a generalization of variance.

Observations:

1. $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$
2. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + 2\mathrm{Cov}(X, Y) + \mathrm{Var}(Y)$
3. $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$

Joint distributions

Conclusion $E(XY)$ can be interpreted as a measure of correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$

Covariance

Let X, Y be random variables with means μ_X, μ_Y respectively. Then

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))$$

Note: If $X = Y$ then $\text{Cov}(X, Y) = \text{Var}(X)$. Hence, covariance is a generalization of variance.
Observations:

1. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
2. $\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$
3. $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$
4. $\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X, Y)$

Joint distributions

Correlation coefficient

Let X, Y be random variables. Then the correlation coefficient is defined by

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(x)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Joint distributions

Correlation coefficient

Let X, Y be random variables. Then the correlation coefficient is defined by

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Observations:

1. $-1 \leq \rho \leq 1$

Joint distributions

Correlation coefficient

Let X, Y be random variables. Then the correlation coefficient is defined by

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(x)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Observations:

1. $-1 \leq \rho \leq 1$
2. if $X = Y$ i.e. fully correlated then $\rho = 1$

Joint distributions

Correlation coefficient

Let X, Y be random variables. Then the correlation coefficient is defined by

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(x)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Observations:

1. $-1 \leq \rho \leq 1$
2. if $X = Y$ i.e. fully correlated then $\rho = 1$
3. if $X = -Y$ i.e. negatively correlated then $= -1$

Joint distributions

Correlation coefficient

Let X, Y be random variables. Then the correlation coefficient is defined by

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(x)\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Observations:

1. $-1 \leq \rho \leq 1$
2. if $X = Y$ i.e. fully correlated then $\rho = 1$
3. if $X = -Y$ i.e. negatively correlated then $= -1$
4. if X and Y are uncorrelated then $\rho = 0$

Joint distributions

Correlation coefficient

Let X, Y be random variables. Then the correlation coefficient is defined by

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(x)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Observations:

1. $-1 \leq \rho \leq 1$
2. if $X = Y$ i.e. fully correlated then $\rho = 1$
3. if $X = -Y$ i.e. negatively correlated then $= -1$
4. if X and Y are uncorrelated then $\rho = 0$
5. if X and Y are independent then $E(XY) = E(X)E(Y)$ and hence $\text{Cov}(X, Y) = 0$ and $\text{Var}(XY) = \text{Var}(X)\text{Var}(Y)$

Joint distributions

Interesting result Let $X^* = \frac{X-\mu_X}{\sigma_X}$ and $Y^* = \frac{Y-\mu_Y}{\sigma_Y}$.

Joint distributions

Interesting result Let $X^* = \frac{X-\mu_X}{\sigma_X}$ and $Y^* = \frac{Y-\mu_Y}{\sigma_Y}$. Then

$$\rho^* == \frac{\text{Cov}(X^*, Y^*)}{\sigma_{X^*}\sigma_{Y^*}} == \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \rho$$

Joint distributions

Interesting result Let $X^* = \frac{X-\mu_X}{\sigma_X}$ and $Y^* = \frac{Y-\mu_Y}{\sigma_Y}$. Then

$$\rho^* == \frac{\mathrm{Cov}(X^*, Y^*)}{\sigma_{X^*}\sigma_{Y^*}} == \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \rho$$

Remark Independent random variables means $\mathrm{Cov}(X, Y) = 0$ i.e. uncorrelated random variable, however the converse need not be true.

Joint distributions

Interesting result Let $X^* = \frac{X-\mu_X}{\sigma_X}$ and $Y^* = \frac{Y-\mu_Y}{\sigma_Y}$. Then

$$\rho^* == \frac{\text{Cov}(X^*, Y^*)}{\sigma_{X^*}\sigma_{Y^*}} == \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \rho$$

Remark Independent random variables means $\text{Cov}(X, Y) = 0$ i.e. uncorrelated random variable, however the converse need not be true.
Counterexample Let Z be a random variable with $\Omega_Z = \{0, 1, 2, 3\}$ and pmf $f(z) = \frac{1}{4}$, $z \in \Omega_Z$. Now define

$$X = \cos\frac{\pi}{2}Z, \; Y = \sin\frac{\pi}{2}Z.$$

Then check that $\text{Cov}(X, Y) = 0$ but $f(x, y) \neq f_X(x)f_Y(y)$. (Homework)

Joint distributions

Joint Moment Generating Function

Let X, Y be random variables with joint pdf $f(x, y)$. Then the real values function

$$M(s, t) = E(e^{sX+tY})$$

is called the joint mgf of X and Y, if it exists in some interval $(s, t) \in [-h, h] \times [-k, k]$

Joint distributions

Joint Moment Generating Function

Let X, Y be random variables with joint pdf $f(x, y)$. Then the real values function

$$M(s, t) = E(e^{sX+tY})$$

is called the joint mgf of X and Y, if it exists in some interval $(s, t) \in [-h, h] \times [-k, k]$

Observations

1. $M(s, 0) = E(e^{sX})$

Joint distributions

Joint Moment Generating Function

Let X, Y be random variables with joint pdf $f(x, y)$. Then the real values function

$$M(s, t) = E(e^{sX+tY})$$

is called the joint mgf of X and Y, if it exists in some interval $(s, t) \in [-h, h] \times [-k, k]$

Observations

1. $M(s, 0) = E(e^{sX})$
2. $M(0, t) = E(e^{ty})$

Joint distributions

Joint Moment Generating Function

Let X, Y be random variables with joint pdf $f(x, y)$. Then the real values function

$$M(s,t) = E(e^{sX+tY})$$

is called the joint mgf of X and Y, if it exists in some interval $(s,t) \in [-h,h] \times [-k,k]$

Observations

1. $M(s,0) = E(e^{sX})$
2. $M(0,t) = E(e^{ty})$
3. $E(X^k) = \frac{\partial^k M(s,t)}{\partial s^k}\Big|_{(0,0)}$, $E(Y^k) = \frac{\partial^k M(s,t)}{\partial t^k}\Big|_{(0,0)}$, $k = 1, 2, 3, \ldots$

Joint distributions

Joint Moment Generating Function

Let X, Y be random variables with joint pdf $f(x, y)$. Then the real values function

$$M(s, t) = E(e^{sX+tY})$$

is called the joint mgf of X and Y, if it exists in some interval $(s, t) \in [-h, h] \times [-k, k]$

Observations

1. $M(s, 0) = E(e^{sX})$
2. $M(0, t) = E(e^{ty})$
3. $E(X^k) = \frac{\partial^k M(s,t)}{\partial s^k}\Big|_{(0,0)}, \; E(Y^k) = \frac{\partial^k M(s,t)}{\partial t^k}\Big|_{(0,0)}, \; k = 1, 2, 3, \ldots$
4. $E(XY) = \frac{\partial^2 M(s,t)}{\partial s \partial t}\Big|_{(0,0)}$

Joint distributions

Joint Moment Generating Function

Let X, Y be random variables with joint pdf $f(x,y)$. Then the real values function

$$M(s,t) = E(e^{sX+tY})$$

is called the joint mgf of X and Y, if it exists in some interval $(s,t) \in [-h,h] \times [-k,k]$

Observations

1. $M(s,0) = E(e^{sX})$
2. $M(0,t) = E(e^{ty})$
3. $E(X^k) = \left.\frac{\partial^k M(s,t)}{\partial s^k}\right|_{(0,0)}, \; E(Y^k) = \left.\frac{\partial^k M(s,t)}{\partial t^k}\right|_{(0,0)}, \; k = 1, 2, 3, \ldots$
4. $E(XY) = \left.\frac{\partial^2 M(s,t)}{\partial s \partial t}\right|_{(0,0)}$
5. if X, Y are independent then $M_{aX+bY}(t) = M_X(at) + M_Y(bt)$

Joint distributions

Conditional pmf/pdf

Let X and Y be discrete random variables. Then the conditional pmf of X given Y is

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

Joint distributions

Conditional pmf/pdf

Let X and Y be discrete random variables. Then the conditional pmf of X given Y is

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

Question What is the interpretation behind this definition!

Joint distributions

Conditional pmf/pdf

Let X and Y be discrete random variables. Then the conditional pmf of X given Y is

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

Question What is the interpretation behind this definition!

Probability of an (conditional) event

Let X, Y be discrete and A be an event. Then

$$P(X \in A | Y = y) = \sum_{x \in A} f_{X|Y}(x|y)$$

and

$$P(X \in A) = \sum_{x \in A} \sum_{y \in \Omega_Y} f_{X|Y}(x|y) f_Y(y) = \sum_{y \in \Omega_Y} P(X \in A | Y = y) f_Y(y)$$

Joint distributions

Probability of conditional event

let X, Y be continuous r.v. and A be an event. Then

$$P(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx$$

and

$$P(X \in A) = \int_{\Omega_Y} P(X \in A|Y = y) f_Y(y)dy$$

Joint distributions

Probability of conditional event

let X, Y be continuous r.v. and A be an event. Then

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y) dx$$

and

$$P(X \in A) = \int_{\Omega_Y} P(X \in A | Y = y) f_Y(y) dy$$

Conditional cdf

Let X, Y be rvs. Then

$$F_{X|Y}(x|y) = P(X \leq x | Y = y) = \begin{cases} \sum_{x' \leq x} f_{X|Y}(x'|y) \text{ if } X, Y \text{ are discrete} \\ \frac{\int_{-\infty}^{x} f(x', y) dx'}{f_Y(y)} \text{ if } X, Y \text{ are continuous} \end{cases}$$

Joint distributions

Conditional expectation

The conditional expectation of X given $Y = y$ is

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

for discrete random variables, and

$$E(X|Y=y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

for continuous random variables

Joint distributions

Conditional expectation

The conditional expectation of X given $Y = y$ is

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

for discrete random variables, and

$$E(X|Y = y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

for continuous random variables

The law of total expectation

$$E(X) = \sum_y E(X|Y = y)\, p_Y(y) \text{ or } E(X) = \int_{-\infty}^{\infty} E(X|Y = y)\, f_Y(y) dy$$

Joint distributions

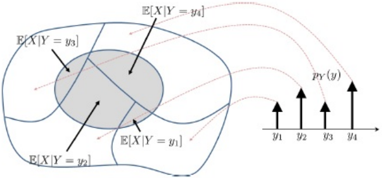
Proof:

$$\begin{aligned} E(X) &= \sum_x x f_X(x) = \sum_x x \left(\sum_y f(x,y) \right) \\ &= \sum_x \sum_y x f_{X|Y}(x|y) f_Y(y) \\ &= \sum_y \left(\sum_x s f_{X|Y}(x|y) \right) f_Y(y) \\ &= \sum_y E(X|Y=y) f_Y(y) \end{aligned}$$

Joint distributions

Proof:

$$\begin{aligned}
E(X) &= \sum_x x f_X(x) = \sum_x x \left(\sum_y f(x,y) \right) \\
&= \sum_x \sum_y x f_{X|Y}(x|y) f_Y(y) \\
&= \sum_y \left(\sum_x s f_{X|Y}(x|y) \right) f_Y(y) \\
&= \sum_y E(X|Y = y) f_Y(y)
\end{aligned}$$



Joint distributions

Note

1. The mean of a random variable X is a deterministic number.

Joint distributions

Note

1. The mean of a random variable X is a deterministic number.
2. The conditional mean of X given $Y = y$, that is, $E(X|y)$, is a function $h(y)$ of y.

Joint distributions

Note

1. The mean of a random variable X is a deterministic number.
2. The conditional mean of X given $Y = y$, that is, $E(X|y)$, is a function $h(y)$ of y.
3. Form the function $h(Y)$, which is a random variable

Joint distributions

Note

1. The mean of a random variable X is a deterministic number.
2. The conditional mean of X given $Y = y$, that is, $E(X|y)$, is a function $h(y)$ of y.
3. Form the function $h(Y)$, which is a random variable

Thus:

$$E(X) = E_Y(E_{X|Y}(X|Y))$$

Joint distributions

Note

1. The mean of a random variable X is a deterministic number.
2. The conditional mean of X given $Y = y$, that is, $E(X|y)$, is a function $h(y)$ of y.
3. Form the function $h(Y)$, which is a random variable

Thus:

$$E(X) = E_Y(E_{X|Y}(X|Y))$$

Proof.

$$E(X) = \sum_y E(X|Y = y) f_Y(y) = \sum_y h(y) f_Y(y) = E(h(Y)) = E(E(X|Y))$$

Joint distributions

Note

1. The mean of a random variable X is a deterministic number.
2. The conditional mean of X given $Y = y$, that is, $E(X|y)$, is a function $h(y)$ of y.
3. Form the function $h(Y)$, which is a random variable

Thus:

$$E(X) = E_Y(E_{X|Y}(X|Y))$$

Proof.

$$E(X) = \sum_y E(X|Y = y) f_Y(y) = \sum_y h(y) f_Y(y) = E(h(Y)) = E(E(X|Y))$$

Remark All the above results can be obtained when $X = x$ is given and Y is free

Joint distributions

A special result Let X and Y be random variables with mean μ_X and μ_Y, and standard deviation σ_X and σ_y, respectively. If conditional expectation of Y given $X = x$ is linear in x, then

$$E(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

Joint distributions

A special result Let X and Y be random variables with mean μ_X and μ_Y, and standard deviation σ_X and σ_y, respectively. If conditional expectation of Y given $X = x$ is linear in x, then

$$E(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

Proof. Assume that X and Y are continuous. Let $E(Y|X = x) = ax + b$. Thus

$$\int_{-\infty}^{\infty} yf_{Y|X}(y|x) = ax + b$$

Joint distributions

A special result Let X and Y be random variables with mean μ_X and μ_Y, and standard deviation σ_X and σ_y, respectively. If conditional expectation of Y given $X = x$ is linear in x, then

$$E(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

Proof. Assume that X and Y are continuous. Let $E(Y|X = x) = ax + b$. Thus

$$\int_{-\infty}^{\infty} y f_{Y|X}(y|x) = ax + b$$

$$\Rightarrow \int_{-\infty}^{\infty} y\frac{f(x,y)}{f_X(x)}dy = ax + b$$

Joint distributions

A special result Let X and Y be random variables with mean μ_X and μ_Y, and standard deviation σ_X and σ_y, respectively. If conditional expectation of Y given $X = x$ is linear in x, then

$$E(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

Proof. Assume that X and Y are continuous. Let $E(Y|X = x) = ax + b$. Thus

$$\begin{aligned} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) &= ax + b \\ \Rightarrow \int_{-\infty}^{\infty} y \frac{f(x,y)}{f_X(x)} dy &= ax + b \\ \Rightarrow \int_{-\infty}^{\infty} y f(x,y) dy &= (ax + b) f_X(x) \\ \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y) dy dx &= \int_{-\infty}^{\infty} (ax + b) f_X(x) dx \end{aligned}$$

Joint distributions

Thus it follows that

$$\mu_Y = a\mu_X + b.$$

Joint distributions

Thus it follows that

$$\mu_Y = a\mu_X + b.$$

Further,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dydx = \int_{-\infty}^{\infty}(ax^2 + bx)f_X(x)dx$$
$$\Rightarrow E(XY) = aE(X^2) + b\mu_X.$$

Joint distributions

Thus it follows that

$$\mu_Y = a\mu_X + b.$$

Further,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dydx = \int_{-\infty}^{\infty}(ax^2+bx)f_X(x)dx$$
$$\Rightarrow E(XY) = aE(X^2) + b\mu_X.$$

Therefore,

$$\begin{aligned} a &= \frac{E(XY)-\mu_X\mu_Y}{\sigma_X^2} \\ &= \frac{\text{Cov}(X,Y)}{\sigma_X^2} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_X}\frac{\sigma_Y}{\sigma_X} \\ &= \rho\frac{\sigma_Y}{\sigma_X} \end{aligned}$$

Joint distributions

Thus it follows that

$$\mu_Y = a\mu_X + b.$$

Further,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dydx = \int_{-\infty}^{\infty}(ax^2+bx)f_X(x)dx$$
$$\Rightarrow E(XY) = aE(X^2) + b\mu_X.$$

Therefore,

$$\begin{aligned} a &= \frac{E(XY)-\mu_X\mu_Y}{\sigma_X^2} \\ &= \frac{\mathrm{Cov}(X,Y)}{\sigma_X^2} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_X}\frac{\sigma_Y}{\sigma_X} \\ &= \rho\frac{\sigma_Y}{\sigma_X} \end{aligned}$$

Similarly, $b = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X$

Joint distributions

Conditional variance

Let X, Y be random variables with joint pdf $f(x, y)$. Let $f_{Y|X}(y|x)$ be the conditional density of Y given $X = x$. Then the conditional variance of Y given $X = x$ is defined as

$$\text{Var}(Y|X = x) = E(Y^2|X = x) - (E(Y|X = x))^2$$

Joint distributions

Conditional variance

Let X, Y be random variables with joint pdf $f(x, y)$. Let $f_{Y|X}(y|x)$ be the conditional density of Y given $X = x$. Then the conditional variance of Y given $X = x$ is defined as

$$\text{Var}(Y|X = x) = E(Y^2|X = x) - (E(Y|X = x))^2$$

An interesting result If X, Y are random variables with mean μ_X, μ_Y, and standard deviation σ_X, σ_Y respectively then

$$E_X(\text{Var}(Y|X)) = (1 - \rho^2)\text{Var}(Y)$$