Lecture 6: Multiple Random Variables

STATS 101: Foundations of Statistics

Linh Tran
linh@thetahat.ai

January 8, 2019

Announcements

- Video links added on course website.
- Top HW scores have been updated.
- Next assignment is posted (due 1/15 @ 9:00am)
  - Partner pairing happens now.
- A *Colab* script is available for today's class.

Outline

Multiple random variables

- Random variable review
- Multiple random variables
- Conditional moments
- Covariance/correlation
- Transformations

Sample space

The set of all possible values is called the *sample space* S.

- It's the space where realizations can be produced.

Examples:

- Tossing a coin

$$S = \{Heads, Tails\} \tag{1}$$

- Rolling a die

$$S = \{1, 2, 3, 4, 5, 6\} \tag{2}$$

- Picking a card

$$S = \{Ace, 2, \ldots, King\} \times \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\} \tag{3}$$

Probability function

A *probability function* is a function $P : \mathcal{B} \to [0, 1]$, where

- $P(\emptyset) = 0$
- $P(S) = 1$
- $P(B_i) \geq 0$
- $P\left(\cup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i)$ when $B_1, B_2, \ldots$ are disjoint

Examples:

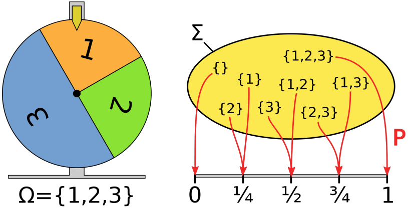
- Tossing a coin
$$S = \{Heads, Tails\} \tag{4}$$
- Rolling a die
$$S = \{1, 2, 3, 4, 5, 6\} \tag{5}$$
- Picking a card
$$S = \{Ace, 2, \ldots, King\} \times \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\} \tag{6}$$

Probability space

Def:
A *probability space* is a triple $(S, \mathcal{B}, P)$.

- S is the set of possible singleton events
- $\mathcal{B}$ is the set of questions to ask P
- P maps sets into probabilities

n.b. They represent the ingredients needed to talk about probabilities



Conditional probability

For events A and B where $P(B) > 0$, the *conditional probability* of A given B (denoted $P(A|B)$) is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{7}$$

Example: In an agricultural region with 1000 farms, we want to know if the farm has vineyards or cork trees.

		Cork Trees	
		Yes	No
Vineyard	Yes	200	50
	No	150	600

Table: Frequency counts

Conditional probability

Example: In an agricultural region with 1000 farms, we want to know if the farm has vineyards or cork trees.

		Cork Trees	
		Yes	No
Vineyard	Yes	20%	5%
	No	15%	60%

Table: Joint probabilities

Questions:

- What is the probability of seeing cork trees in a farm with vineyards?
- Among farms with cork trees or vineyards, what is the probability of having both?

Conditional probability

Let's assume the following joint probabilties

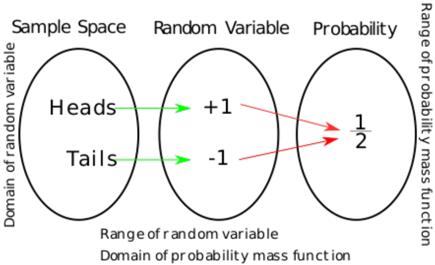
		Cork Trees	
		Yes	No
Vineyard	Yes	25%	25%
	No	25%	25%

We have that $P(A \cap B) = P(A) \cdot P(B)$, meaning that they are *independent*

Random variables

A *random variable* is a (Borel measureable) function
$X : S \to \mathbb{R}$
Example: For coin tossing, we have $X : \{Heads, Tails\} \to \mathbb{R}$, where

$$X(s) = \begin{cases} 1 & \text{if } s = Heads \\ 0 & \text{if } s = Tails \end{cases} \tag{8}$$



Cumulative distribution function

The *cumulative distribution function* (cdf) of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$.

Cumulative distribution function

The *cumulative distribution function* (cdf) of a random variable X is the function $F_X : \mathbb{R} \to [0,1]$.

Example: For coin tossing, we have
$X : \{Heads, Tails\} \to \mathbb{R}$,

where

$$X(s) = \begin{cases} 1 & \text{if } s = Heads \\ 0 & \text{if } s = Tails \end{cases} \tag{9}$$

we have

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} \tag{10}$$

Cumulative distribution function

The *cumulative distribution function* (cdf) of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$.

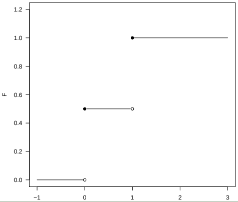
Example: For coin tossing, we have
$X : \{Heads, Tails\} \to \mathbb{R}$,

where

$$X(s) = \begin{cases} 1 & \text{if } s = Heads \\ 0 & \text{if } s = Tails \end{cases} \tag{9}$$

we have

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases} \tag{10}$$



Distributions

Most useful distributions have names, e.g.

- Normal distribution
- Uniform distribution
- Bernoulli distribution
- Binomial distribution
- Poisson distribution
- Gamma distribution

Colab link

Multiple random variables

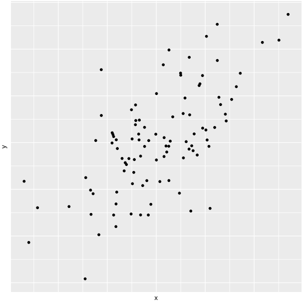
A n dimensional vector $\mathbf{X} = (X_1, ..., X_n)'$ is a random vector if $X_1, ..., X_n$ are rv's defined on the same probability space.

Multiple random variables

A n dimensional vector $\mathbf{X} = (X_1, ..., X_n)'$ is a random vector if $X_1, ..., X_n$ are rv's defined on the same probability space.

The joint cdf of a vector (X, Y) is $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$,

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \; \forall (x, y)' \in \mathbb{R}^2 \tag{11}$$



Multiple random variables

The joint pmf/pdf of (X, Y):

- $(X, Y)'$ is *discrete* if $\exists f_{X,Y} : \mathbb{R}^2 \to [0, 1]$ such that

$$F_{X,Y}(x, y) = \sum_{s \leq x} \sum_{t \leq y} f_{X,Y}(s, t) \ \forall (x, y)' \in \mathbb{R}^2 \qquad (12)$$

- $(X, Y)'$ is *continuous* if $\exists f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_+$ such that

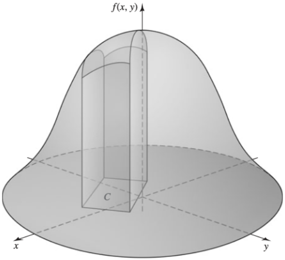
$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{\infty}^{y} f_{X,Y}(s, t) dtds \ \forall (x, y)' \in \mathbb{R}^2 \qquad (13)$$

- $(X, Y)'$ is *mixed* if $\exists f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}_+$ such that

$$F_{X,Y}(x, y) = \sum_{s \leq x} \int_{\infty}^{y} f_{X,Y}(s, t) dt \ \forall (x, y)' \in \mathbb{R}^2 \qquad (14)$$

Example: A bivariate normal distribution

Figure 3.11 An example of a joint p.d.f.



Colab link

Multiple random variables

Note: all properties we've covered for univariate distributions can be extended to multivariate distributions, e.g. Let $g(x, y) : \mathbb{R}^2 \to \mathbb{R}$. Then

$$\mathbb{E}[g(x,y)] = \begin{cases} \sum_{x,t \in \mathbb{R}^2} g(s,t) f_{X,Y}(s,t) & \text{if } (x,y) \text{ is discrete} \\ \int_{\infty}^{\infty} \int_{\infty}^{\infty} g(s,t) f_{X,Y}(s,t) dtds & \text{if } (x,y) \text{ is continuous} \end{cases} \tag{15}$$

Marginal distributions

For bivariate random vector $(X, Y)'$, the cdf of X (and of Y) is called the *marginal cdf* of X(and of Y), e.g.

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) \; \forall x \in \mathbb{R} \tag{16}$$

Marginal distributions

For bivariate random vector $(X, Y)'$, the cdf of X (and of Y) is called the *marginal cdf* of X(and of Y), e.g.

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) \; \forall x \in \mathbb{R} \tag{16}$$

We can also obtain pmf/pdf the same way, i.e.

- If $(X, Y)'$ is discrete, then

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y) \; x \in \mathbb{R} \tag{17}$$

- If $(X, Y)'$ is continuous, then

$$f_X(x) = \int_{\infty}^{\infty} f_{X,Y}(x, y) dy \; x \in \mathbb{R} \tag{18}$$

Marginal distributions

Example: Let X, Y have the following joint pmf/pdf:

$$f(x, y) = \frac{xy^{x-1}}{3} : x = 1, 2, 3; 0 < y < 1 \tag{19}$$

We obtain the *marginal pmf of X* by integrating:

$$f_X(x) = \int_0^1 \frac{xy^{x-1}}{3} dy = \frac{y^x}{3}\bigg|_0^1 = \frac{1}{3} \tag{20}$$

We obtain the *marginal pdf of Y* by summing:

$$f_Y(y) = \sum_x \frac{xy^{x-1}}{3} = \frac{1}{3} + \frac{2y}{3} + y^2 : 0 < y < 1 \tag{21}$$

Marginal distributions

Two random variables X and Y are independent if

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \; \forall (x,y)' \in \mathbb{R}^2 \tag{22}$$

Equivalently, X and Y are independent if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \; \forall (x,y)' \in \mathbb{R}^2 \tag{23}$$

n.b. Knowing $F_{X,Y}(x,y)$ implies knowledge of the marginal distributions. The converse only holds true if $X \perp\!\!\!\perp Y$.

Conditional distributions

The *conditional pmf/pdf* of Y given $X = x$ is given by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \tag{24}$$

Notes:

- If $f_X(x) > 0$, then all properties of pmfs/pdfs apply to the conditional pmfs/pdfs.
- We can interpret this as $P(Y = y)$ given that $X = x$.
- In the continuous case, we always have that $P(X = x) = 0$.

Conditional distributions

Example: Let X, Y have the following joint pmf/pdf:

$$f(x,y) = \frac{xy^{x-1}}{3} : x = 1,2,3; 0 < y < 1 \tag{25}$$

We obtain the *conditional pdf of Y given X* by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{xy^{x-1}}{3}}{\frac{1}{3}} = xy^{x-1} \tag{26}$$

We obtain the *conditional pmf of X given Y* by

$$f_{X|Y}(x|y) = \frac{\frac{xy^{x-1}}{3}}{\frac{1}{3} + \frac{2y}{3} + y^2} = \frac{xy^{x-1}}{1 + 2y + 3y^2} \tag{27}$$

Law of iterated expectations

For random vector $(X, Y)'$

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] \tag{28}$$

provided that the expectations exist.

Law of iterated expectations

For random vector $(X, Y)'$

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] \tag{28}$$

provided that the expectations exist.

Proof:

$$\begin{aligned}
\mathbb{E}[Y] &= \int\int y f_{X,Y}(x,y)dxdy &(29)\\
&= \int\int y f_{Y|X}(y|x) f_X(x)dxdy &(30)\\
&= \int\int y f_{Y|X}(y|x)dy f_X(x)dx &(31)\\
&= \mathbb{E}_X[\mathbb{E}_{Y|X}[Y|X]] &(32)
\end{aligned}$$

Covariance

The covariance of X and Y is

$$Cov(X, Y) = \sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \tag{33}$$

Properties:

- $Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $Cov(X, X) = var(X)$
- $Cov(X, Y) = 0$ if $X \perp\!\!\!\perp Y$

Covariance

Example: Let X, Y be continuous rv's such that

$$f(x, y) = x + y : 0 \leq x, y \leq 1 \tag{34}$$

To get the covariance, we: (i) get the marginal distributions, (ii) get the expectations, and (iii) use the covariance formula.

Covariance

Example: Let X, Y be continuous rv's such that

$$f(x,y) = x + y : 0 \leq x, y \leq 1 \tag{34}$$

To get the covariance, we: (i) get the marginal distributions, (ii) get the expectations, and (iii) use the covariance formula.

$$f(x) = \int_0^1 x + y dy = x + \frac{1}{2},\ f(y) = \int_0^1 x + y dx = y + \frac{1}{2} \tag{35}$$

$$\mathbb{E}[X] = \int_0^1 xf(x)dx = 7/12,\ \mathbb{E}[Y] = \int_0^1 yf(y)dy = 7/12 \tag{36}$$

$$\begin{aligned} Cov(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] & (37)\\ &= \int_0^1 \int_0^1 (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) f(x,y) dxdy & (38)\\ &= \int_0^1 \int_0^1 \left(x - \frac{7}{12}\right)\left(y - \frac{7}{12}\right)(x+y)dxdy & (39)\\ &= -1/144 & (40) \end{aligned}$$

Correlation

The correlation of X and Y is

$$Corr(X,Y) = \rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{41}$$

n.b. The correlation is just $Cov(X,Y)$ standardized by the product of the individual standard deviations.

Properties:

- ρ_{XY} measures linear dependence.
- If $X \perp\!\!\!\perp Y$, then $\rho_{XY} = 0$.
- $|\rho_{XY}| \leq 1$, by the Cauchy-Schwarz inequality.
- $|\rho_{XY}| = 1$ if $P(Y = aX \pm b) = 1$ for some $a \neq 0, b \in \mathbb{R}$.

Correlation

Example: Let X, Y be continuous rv's such that

$$f(x,y) = x + y : 0 \leq x, y \leq 1 \tag{42}$$

Recall: we've already calculated the covariance. To get the correlation, we just need $\sigma_X \sigma_Y$.

$$\begin{aligned}
\sigma_X^2 &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_0^1 (x - 7/12)\left(x + \frac{1}{2}\right) dx && (43)\\
&= 11/144 && (44)\\
\sigma_Y^2 &= \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \int_0^1 (y - 7/12)\left(y + \frac{1}{2}\right) dy && (45)\\
&= 11/144 && (46)
\end{aligned}$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-1/144}{\sqrt{11/144} \cdot \sqrt{11/144}} = -1/11 \tag{47}$$

Transformations

Let $X_1, ..., X_n \sim f_0$ and consider new random variables $Y_1, ..., Y_n$ be generated as:

$$\begin{aligned} Y_1 &= r_1(X_1, ..., X_n) \\ Y_2 &= r_2(X_1, ..., X_n) \\ &\cdots \\ Y_n &= r_n(X_1, ..., X_n) \end{aligned}$$

such that the transformations are one-to-one. Then, for $s_i : x_i = s_i(y_1, ..., y_n)$ the Jacobian of the transformation is

$$J = \det \begin{bmatrix} \frac{\partial s_1}{\partial y_1} & \cdots & \frac{\partial s_1}{\partial y_n} \\ \cdots & \ddots & \cdots \\ \frac{\partial s_n}{\partial y_1} & \cdots & \frac{\partial s_n}{\partial y_n} \end{bmatrix} \tag{48}$$

and the joint pdf of $Y_1, ..., Y_n$ is

$$g(y_1, ..., y_n) = f(s_1, ..., s_n)|J| \tag{49}$$

Transformations

Example: Let X_1, X_2 be rv's such that

$$f(x_1, x_2) = \begin{cases} 4x_1 x_2 & \text{for } 0 < x_1, x_2 < 1 \\ 0 & \text{otherwise} \end{cases} \tag{50}$$

Let $Y_1 = \frac{X_1}{X_2}$, $Y_2 = X_1 X_2$. What is the joint pdf $g(y_1, y_2)$?

Let $Y_1 = \frac{X_1}{X_2}$, $Y_2 = X_1 X_2$. What is the joint pdf $g(y_1, y_2)$?

Inverting the transformation gives us:
$x_1 = (y_1 y_2)^{1/2}, x_2 = \left(\frac{y_2}{y_1}\right)^{1/2}$.

The Jacobian of the transformation is

$$J = \det \begin{bmatrix} \frac{1}{2}\left(\frac{y_2}{y_1}\right)^{1/2} & \frac{1}{2}\left(\frac{y_1}{y_2}\right)^{1/2} \\ -\frac{1}{2}\left(\frac{y_2}{y_1^3}\right)^{1/2} & \frac{1}{2}\left(\frac{1}{y_1 y_2}\right)^{1/2} \end{bmatrix} = \frac{1}{2y_1} \tag{51}$$

Thus, the joint pdf $g(y_1, y_2)$ is

$$\begin{aligned} g(y_1, y_2) &= f((y_1 y_2)^{1/2}, \left(\frac{y_2}{y_1}\right)^{1/2})|J| \\ &= 2\left(\frac{y_2}{y_1}\right) \end{aligned}$$

References

- DeGroot & Schervish Chapters 3, 4, 5