Lecture note 7: Further properties of expectation

Foundation of Probability Theory/STA 203

Zhuosong ZHANG

Department of Statistics and Data Science, SUSTech

Fall, 2023

Expectation of sums of random variables

Function of two random variables

- We have known that, for a discrete random variable X and a function g,

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{S}} g(x) p(x).$$

- How about $g(X, Y)$?

Definition 1

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, on which there are two random variables X and Y. Let $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, then $g(X, Y)$ is a random variable such that

$$g(X, Y)(\omega) = g(X(\omega), Y(\omega)) \quad \text{for } \omega \in \Omega.$$

Proposition 2

If discrete random variables X and Y have a joint probability mass function $p(x,y)$, then for $g:\mathbb{R}^2 \longrightarrow \mathbb{R}$,

$$\mathbb{E}[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y).$$

Example 3

Let $\Omega_1 = \{2,3,4,5\}$, and let $\Omega_2 = \{2,3,4,5\}$, and let $\Omega = \Omega_1 \times \Omega_2 = \{(2,2),(2,3),\dots,(5,5)\}$. Let X be the value of the first number and Y be that of the second number. We have shown that X and Y are independent. Let $g(x,y) = x+y$. What is the pmf of $g(X,Y)$?

Proof.

Write $Z = g(X, Y) = X + Y$. Note that

$$\mathbb{P}\{Z = 4\} = \mathbb{P}\{X = 2, Y = 2\} = \mathbb{P}\{(2,2)\} = \frac{1}{16},$$

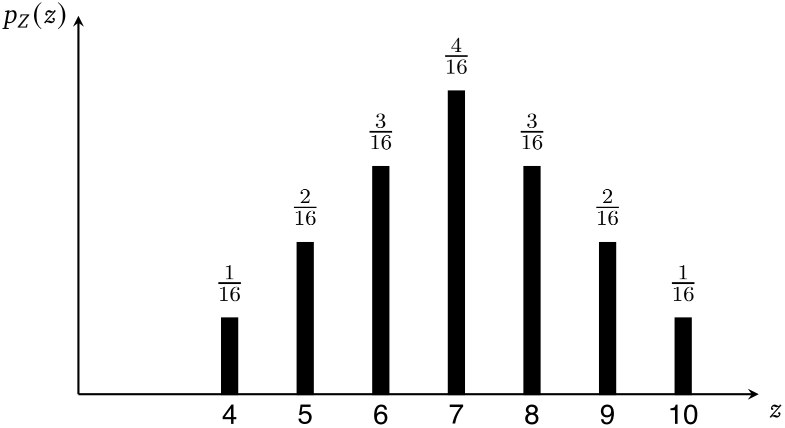
$$\begin{aligned}\mathbb{P}\{Z = 5\} &= \mathbb{P}\{X = 2, Y = 3\} + \mathbb{P}\{X = 3, Y = 2\} \\ &= \mathbb{P}\{(2,3),(3,2)\} = \frac{1}{8},\end{aligned}$$

$$\begin{aligned}\mathbb{P}\{Z = 6\} &= \mathbb{P}\{X = 2, Y = 4\} + \mathbb{P}\{X = 3, Y = 3\} + \mathbb{P}\{X = 4, Y = 2\} \\ &= \mathbb{P}\{(2,4),(3,3),(4,2)\} = \frac{3}{16},\end{aligned}$$

$$\vdots$$

$$\mathbb{P}\{Z = 10\} = \mathbb{P}\{X = 5, Y = 5\} = \mathbb{P}\{(5,5)\} = \frac{1}{16}.$$

■



$p_Z(z)$
$\frac{1}{16}$
$\frac{2}{16}$
$\frac{3}{16}$
$\frac{4}{16}$
$\frac{3}{16}$
$\frac{2}{16}$
$\frac{1}{16}$
4
5
6
7
8
9
10
z

Expectation of $g(X, Y)$

Proposition 4

If X and Y are jointly continuous with pdf $f(x, y)$, then

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy.$$

Proof.

We only give a proof when X and Y are continuous. Note that

$$\mathbb{E}[g(X,Y)] = \int_0^\infty \mathbb{P}\{g(X,Y) > t\}dt - \int_{-\infty}^0 \mathbb{P}\{g(X,Y) < t\}dt = I_1 - I_2.$$

For the first term I_1,

$$\mathbb{P}\{g(X,Y) > t\} = \iint_{(x,y):g(x,y)>t} f(x,y)dydx,$$

and thus

$$I_1 = \int_0^\infty \iint_{(x,y):g(x,y)>t} f(x,y)dydxdt = \int_{-\infty}^\infty \int_{-\infty}^\infty \int_0^{\max\{g(x,y),0\}} f(x,y)dtdydx$$

$$= \int_{-\infty}^\infty \int_{-\infty}^\infty \max\{g(x,y),0\} f(x,y)dydx.$$

Similarly,

$$I_2 = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \min\{g(x,y),0\}f(x,y)dydx.$$

Therefore,

$$\begin{aligned}\mathbb{E}[g(X,Y)] = I_1 - I_2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\max\{g(x,y),0\} - \min\{g(x,y),0\})f(x,y)dydx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dydx\end{aligned}$$

as desired. ∎

Example 5

An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident, an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

Question

What we need to calculate?

Solution.

We need to compute $\mathbb{E}[|X-Y|]$. Since the joint density function of X and Y is

$$f(x,y) = \frac{1}{L^2} \quad 0 < x, y < L,$$

it follows that

$$\begin{aligned}\mathbb{E}[|X-Y|] &= \frac{1}{L^2}\int_0^L \int_0^L |x-y| dydx \\ &= \frac{L}{3}.\end{aligned}$$

■

Expectation of $\mathbb{E}[X+Y]$

Proposition 6

For any random variables X and Y and real numbers a and b,

$$\mathbb{E}[aX+bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y].$$

More generally,

$$\mathbb{E}[a_1X_1 + \cdots + a_nX_n] = \sum_{i=1}^{n} a_i\,\mathbb{E}[X_i].$$

Monotonicity

Proposition 7

If $X \geqslant Y$ a.s., then

$$\mathbb{E}[X] \geqslant \mathbb{E}[Y].$$

Example 8 (Expectation of a binomial random variable)

Let X be a binomial random variable with parameters n and p. Let $X_1, \ldots, X_n$ be defined as

$$X_i = \begin{cases} 1 & \text{it the } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

Hence, $X = X_1 + \cdots + X_n$. Note that

$$\mathbb{E}[X_i] = p \quad \text{for each } i,$$

then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np.$$

Sample mean

Example 9

Let $X_1, \ldots, X_n$ be i.i.d. random variables having the distribution function F and expected value μ. Such a sequence of random variables is said to be a sample from the distribution F. The quantity

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

is called the sample mean. Compute $\mathbb{E}[\bar{X}]$.

Example 10

Let $A_1, \ldots, A_n$ denote events, and define the indicator variable X_i of A_i by

$$X_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Let

$$X = \sum_{i=1}^{n} X_i$$

be the number of the events A_i that occur, and let

$$Y_i = \begin{cases} 1 & \text{if } X \geqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

be the variable that indicates that at least one of the A_i occurs. Then, by the monotonicity property,

$$\mathbb{E}[Y] \leqslant \mathbb{E}[X],$$

while

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

and

$$\mathbb{E}[Y] = \mathbb{P}\{\text{at least one of } A_i \text{ occur}\}$$
$$= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right).$$

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 11 (Expected number of matches)

Suppose that N people throw their hats into the center of a room. The hats are mixed up, and each person randomly selects one. Find the expected number of people that select their own hat.

Solution.

Let X denote the number of matches, then

$$X = X_1 + \cdots + X_n,$$

where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th person selects his own hat,} \\ 0 & \text{otherwise.} \end{cases}$$

Examples

For each i, the ith person is equally likely to select any of the N hats, then

$$\mathbb{E}[X_i] = \mathbb{P}\{X_i = 1\} = \frac{1}{N}.$$

Thus,

$$\mathbb{E}[X] = \sum_{i=1}^{N} \mathbb{E}[X_i] = 1.$$

■

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 12 (Coupon-collecting problems)

Suppose that there are N different types of coupons, and each time one obtains a coupon, it is equally likely to be any one of the N types. Find the expected number of coupons one need amass before obtaining a complete set of at least one of each type.

Solution.

Let X denote the number of coupons collected before a complete set is attained. Note that

$$X = X_0 + \cdots + X_{N-1},$$

where X_i is the number of additional coupons that need be obtained after i distinct types have been collected in order to obtain another distinct type. Then, a new coupon obtained will be of a distinct type with probability $(N - i)/N$.

Examples

Therefore,

$$\mathbb{P}\{X_i = k\} = \frac{N-i}{N}\left(\frac{i}{N}\right)^{k-1}, \quad \text{or} \quad X_i \sim \text{Geometric}(\frac{N-i}{N}).$$

Therefore, $\mathbb{E}[X_i] = \frac{N}{N-i}$, and hence,

$$\mathbb{E}[X] = \frac{N}{N} + \frac{N}{N-1} + \cdots + \frac{N}{1}$$
$$\to N \log N \quad \text{as } N \to \infty.$$

■

Covariance

Products of independent random variables

Proposition 13

If X and Y are independent, then, for any functions h and g,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)].$$

Proof.

Suppose that X and Y are jointly continuous with density $f(x, y)$. Then,

$$\begin{aligned}\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= \mathbb{E}[h(Y)]\,\mathbb{E}[g(X)].\end{aligned}$$

■

Covariance

Definition 14 (Covariance)

The covariance between X and Y, denoted by $\mathrm{Cov}(X,Y)$ is defined by

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

Proposition 15

If X and Y are independent, then

$$\mathrm{Cov}(X,Y) = 0.$$

The inverse is not always correct.

Example: uncorrelated but not independent

Example 16

Let X and Y be defined as follows:

$$X = \begin{cases} 1 & \text{with probability } 1/3, \\ 0 & \text{with probability } 1/3, \\ -1 & \text{with probability } 1/3, \end{cases}$$

$$Y = X^2.$$

Whether X and Y are independent? Are they correlated?

Example: uncorrelated but not independent

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Solution.

To see whether X and Y are correlated, we calculate the covariance between X and Y. Note that $\mathbb{E}[X] = 0$, $\mathbb{E}[Y] = 1$, and $\mathbb{E}[XY] = \mathbb{E}[X^3] = 0$. Therefore,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0 - (0)(1) = 0.$$

Example: uncorrelated but not independent

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

The joint distribution of X and Y is given by

	$X=-1$	$X=0$	$X=1$
$Y=0$	0	1/3	0
$Y=1$	1/3	0	1/3

Note that

$$\mathbb{P}\{X=-1, Y=0\} = 0,$$

but

$$\mathbb{P}\{X=-1\} = \frac{1}{3}, \quad \mathbb{P}\{Y=0\} = \mathbb{P}\{X=0\} = \frac{1}{3}.$$

Therefore, X is not independent of Y. ■

More properties of covariance

Proposition 17

(i) Symmetry: $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(ii) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

(iii) $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y)$.

(iv) $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.

(v) For $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ and $a_1, \ldots, a_n, b_1, \ldots, b_m$, we have

$$\mathrm{Cov}(a_1X_1 + \cdots + a_nX_n, b_1Y_1 + \cdots + b_mY_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\,\mathrm{Cov}(X_i, Y_j).$$

Examples

Example 18 (Sample mean and sample variance)

Let $X_1, \ldots, X_n$ be a sample from F with mean μ and variance σ^2. Let

$$\bar{X} = (X_1 + \cdots + X_n)/n$$

be the sample mean. The sample variance is defined as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Find $\mathrm{Var}(\bar{X})$ and $\mathbb{E}[S^2]$.

Solution.

By (ii) and (v) of Proposition 17,

$$\mathrm{Var}(\bar{X}) = \mathrm{Cov}(\bar{X}, \bar{X}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{n^2} \mathrm{Cov}(X_i, X_j) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_i) = \frac{\sigma^2}{n}.$$

For $\mathbb{E}[S^2]$, we assume $\mu = 0$ without loss of generality, observe that

$$\begin{aligned} \mathbb{E}[(X_i - \bar{X})^2] &= \mathbb{E}[X_i^2] - 2\,\mathbb{E}[X_i \bar{X}] + \mathbb{E}[\bar{X}^2] \\ &= \sigma^2 - \frac{2}{n}\sigma^2 + \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2. \end{aligned}$$

Thus,

$$\mathbb{E}[S^2] = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{n-1}{n} \sigma^2 \right) = \sigma^2. \quad \blacksquare$$

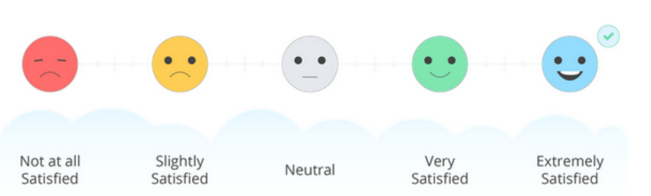
Example 19 (Binomial random variable)

Compute the variance of a binomial random variable X with parameters n and p.

Example 20

Consider a set of N people, each of whom has an opinion about a certain subject that is measured by a real number v that represents the person's "strength of feeling" about the subject. Let v_i represent the strength of feeling of person $i, i = 1, \ldots N$.

Suppose that the quantities $v_i, i = 1, \ldots, N$, are unknown and, to gather information, a group of n of the N people is "randomly chosen without replacement". If S denotes the sum of the n sampled values, determine its mean and variance.



Solution.

For each person i, define an indicator variable I_i to indicate whether or not that person is included in the sample. Then,

$$S = \sum_{i=1}^{N} v_i I_i.$$

Therefore,

$$\mathbb{E}[S] = \sum_{i=1}^{N} v_i \, \mathbb{E}[I_i], \quad \mathrm{Var}(S) = \sum_{i=1}^{N} v_i^2 \, \mathrm{Var}(I_i) + 2 \sum_{i<j} v_i v_j \, \mathrm{Cov}(I_i, I_j).$$

It suffices to find $\mathbb{E}[I_i], \mathrm{Var}(I_i)$ and $\mathrm{Cov}(I_i, I_j)$.

Note that

$$\mathbb{E}[I_i] = \mathbb{P}\{\text{the } i\text{th person is selected}\} = \frac{n}{N},$$

$$\mathbb{E}[I_iI_j] = \mathbb{P}\{\text{both the } i\text{th and } j\text{th are selected}\} = \frac{n}{N}\frac{n-1}{N-1}.$$

Then,

$$\mathrm{Var}(I_i) = \mathbb{E}[I_i^2] - (\mathbb{E}[I_i])^2 = \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n}{N}\left(1 - \frac{n}{N}\right),$$

$$\mathrm{Cov}(I_i, I_j) = \mathbb{E}[I_iI_j] - \mathbb{E}[I_i]\,\mathbb{E}[I_j] = \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2 = -\frac{n(N-n)}{N^2(N-1)}.$$

Substituting these results to the proceeding equation gives the final result. ■

Correlation

Definition 21 (Correlation)

The correlation of two random variables X and Y, denoted by $\text{Cor}(X,Y)$ or $\rho(X,Y)$, is defined as

$$\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}.$$

How to understand correlation?

- Correlation is a statistical measure that quantifies the strength and direction of the linear relationship between two variables.
- The value of ρ ranges between -1 and 1, with a value of 0 indicating no linear relationship, a positive value indicating a positive linear relationship, and a negative value indicating a negative linear relationship.

The Cauchy–Schwarz inequality

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 22

For any random variables X and Y,

$$|\mathbb{E}[XY]| \leqslant \sqrt{\mathbb{E}[X^2]\,\mathbb{E}[Y^2]}.$$

The equality holds if and only if $X = cY$ almost surely, or,

$$\mathbb{P}\{X = cY\} = 1 \quad \text{for some constant } c.$$

Properties of correlation

Proposition 23

For any random variables X and Y,

$$|\operatorname{Cor}(X,Y)| \leqslant 1.$$

Moreover, $\operatorname{Cor}(X,Y) = 1$ if and only if

$$\mathbb{P}\left\{\frac{X-\mathbb{E}[X]}{\operatorname{SD}(X)} = \frac{Y-\mathbb{E}[Y]}{\operatorname{SD}(Y)}\right\} = 1;$$

and $\operatorname{Cor}(X,Y) = -1$ if and only if

$$\mathbb{P}\left\{\frac{X-\mathbb{E}[X]}{\operatorname{SD}(X)} = -\frac{Y-\mathbb{E}[Y]}{\operatorname{SD}(Y)}\right\} = 1.$$

Example

Example 24

If $U \sim U(0, 2\pi)$, $X = \cos(U)$, and $Y = \cos(U + t)$. Find the correlation between X and Y.

Solution

Solution.

Note that

$$\mathbb{E}[X] = \frac{1}{2\pi}\int_0^{2\pi} \cos u du = 0, \quad \mathbb{E}[X^2] = \frac{1}{2\pi}\int_0^{2\pi} \cos^2 u du = \frac{1}{2}.$$

Similarly,

$$\mathbb{E}[Y] = 0, \quad \mathbb{E}[Y^2] = \frac{1}{2}.$$

For the covariance,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - 0 = \frac{1}{2\pi}\int_0^{2\pi} \cos u \cos(u+t) du = \frac{1}{2}\cos t.$$

Solution

Therefore,

$$\rho = \mathrm{Cor}(X, Y) = \cos t.$$

If $t = 0$, then $\rho = 1$, $X = Y$. If $t = \pi$, then $\rho = -1$, $X = -Y$. In both cases, they have a linear relation.

If $t = \pi/2$ or $t = 3\pi/2$, then $\rho = 0$, which means X and Y are uncorrelated. However, since $X^2 + Y^2 = 1$, it follows that X and Y are not independent. ■

Conditional expectation

Definition

Definition 25 (Contiditional expectation)

If X and Y are jointly discrete random variables, then the conditional expectation of X given that $Y = y$, for all values of y such that $p_Y(y) > 0$, by

$$\mathbb{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y).$$

Example 26

If X and Y are independent binomial random variables with identical parameters n and p, calculate $\mathbb{E}[X|X+Y=m]$.

Solution.

We first calculate the conditional pmf of X given that $X+Y=m$:

$$\begin{aligned}
&\mathbb{P}\{X=k|X+Y=m\} \\
&= \frac{\mathbb{P}\{X=k, X+Y=m\}}{\mathbb{P}\{X+Y=m\}} = \frac{\mathbb{P}\{X=k, Y=m-k\}}{\mathbb{P}\{X+Y=m\}} \\
&= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} = \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}.
\end{aligned}$$

Examples

Therefore,

$$\mathbb{E}[X|X+Y=m] = \binom{2n}{m}^{-1} \sum_{k=0}^{\min(n,m)} k\binom{n}{k}\binom{n}{m-k}$$
$$= \frac{m}{2}.$$

∎

Conditional expectation

Definition 27

If X and Y are jointly continuous with PDF $f(x, y)$, then the conditional expectation of X given $Y = y$ is defined by

$$\mathbb{E}[X|Y = y] = \begin{cases} \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx & f_Y(y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Examples

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 28

Suppose that the joint density of X and Y is given by

$$f(x, y) = \frac{e^{-x/y-y}}{y}, \quad 0 < x, y < \infty.$$

Compute $\mathbb{E}[X|Y = y]$.

Examples

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Solution.

The conditional density is, for $y > 0$,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{y^{-1}e^{-x/y-y}}{\int_0^\infty (1/y)e^{-x/y-y}dx}$$

$$= \frac{(1/y)e^{-x/y}}{\int_0^\infty (1/y)e^{-x/y}dx} = \frac{1}{y}e^{-x/y}.$$

Examples

Therefore,

$$\mathbb{E}[X|Y=y] = \int_0^\infty \frac{1}{y} x e^{-x/y} dy$$
$$= y.$$

■

Remark

For the sake of simplicity, if $\mathbb{E}[X|Y=y] = h(y)$ for some function h, then we can write $\mathbb{E}[X|Y] = h(Y)$. Generally, $\mathbb{E}[X|Y]$ is always a function of Y, and hence it is a random variable.

Further properties

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Proposition 29

Let g be a real-valued function, then

$$\mathbb{E}[g(X)|Y=y] = \begin{cases} \sum_x g(x) p_{X|Y}(x|y) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} g(x) p_{X|Y}(x|y) dx & \text{in the continuous case} \end{cases}$$

Moreover, for $X_1, \ldots, X_n$ and real numbers $a_1, \ldots, a_n$,

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i \,\middle|\, Y = y\right] = \sum_{i=1}^{n} \mathbb{E}[X_i|Y=y].$$

Least Squares (最小二乘法)

- Conditional expectation plays an important role in prediction.
- Assume that we have two random variables X and Y.
- For example, assume that we want to know whether there is a relation between the speed and stopping distance in the car dataset.
- Say, Y means the speed of a car, and X means the stopping distance.
- Assume that there is a relation $Y = g(X) + \varepsilon$, where ε means the measurement error, which is independent of X.
- We want to find a function h such that

$$h = \arg\min_{g} \mathbb{E}[\{Y - g(X)\}^2].$$

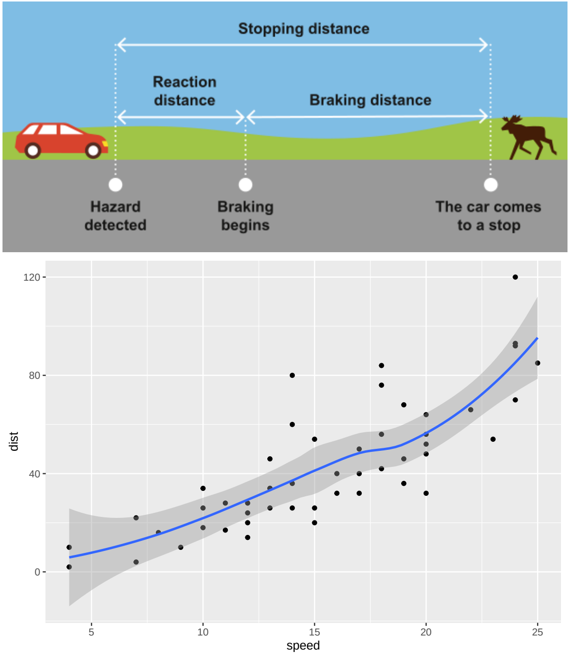


Figure: Speed and stopping distance

Proposition 30

For any random vector (X, Y), let

$$h(x) = \mathbb{E}[Y|X = x].$$

Then, for any function g,

$$\mathbb{E}[(Y - h(X))^2] \leqslant \mathbb{E}[(Y - g(X))^2].$$

In other words, $h(x) = \mathbb{E}[Y|X = x]$ minimizes the loss function $L(g) = \mathbb{E}[(Y - g(X))^2]$.
We call $h(x) = \mathbb{E}[Y|X = x]$ is the regression function.

Total expectation

- Define $\mathbb{E}[X|Y]$ as a random variable which is a function of Y whose value at $Y = y$ is $\mathbb{E}[X|X = y]$.
- For example, if $\mathbb{E}[X|Y = y] = h(y)$, then $\mathbb{E}[X|Y] = h(Y)$.
- An important property of conditional expectation is the following:

Proposition 31 (Total expectation formula)

We have

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|Y]\big] = \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y) dy.$$

Example 32

A miner is trapped in a mine containing 3 doors where there are no differences from the appearance. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

Solution.

Let X denote the amount of time (in hours) until the miner reaches safety, and let Y denote the door he initially chooses. Now,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X|Y=1]\,\mathbb{P}\{Y=1\} + \mathbb{E}[X|Y=2]\,\mathbb{P}\{Y=2\} + \mathbb{E}[X|Y=3]\,\mathbb{P}\{Y=3\} \\ &= \frac{1}{3}\big(\mathbb{E}[X|Y=1] + \mathbb{E}[X|Y=2] + \mathbb{E}[X|Y=3]\big),\end{aligned}$$

and

$$\mathbb{E}[X|Y=1] = 3, \quad \mathbb{E}[X|Y=2] = 5 + \mathbb{E}[X], \quad \mathbb{E}[X|Y=3] = 7 + \mathbb{E}[X].$$

Hence, $\mathbb{E}[X] = 15$. ■

Example 33

Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean of $8. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day?

Solution.

If we let N denote the number of customers that enter the store and X_i the amount spent by the ith such customer, then the total amount of money spent can be expressed as $\sum_{i=1}^{N} X_i$. Now,

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right],$$

and

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n} X_i | N = n\right] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] \qquad \text{(by the independence of } X_i \text{ and } N\text{)} \\ &= n\,\mathbb{E}[X],\end{aligned}$$

which implies that

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}[N\,\mathbb{E}[X]] = \mathbb{E}[N]\,\mathbb{E}[X] = 50 \times 8 = 400. \qquad \blacksquare$$

Examples

Example 34

Independent trials, each resulting in a success with probability p, are successively performed. Let N be the time of the first success. Find $\mathrm{Var}(N)$.

Solution.

To calculate $\mathrm{Var}(N)$, we need to find $\mathbb{E}[N]$ and $\mathbb{E}[N^2]$. We have known that $\mathbb{E}[N] = 1/p$, so it suffices to compute $\mathbb{E}[N^2]$.
Let $Y = 1$ if the first trial is a success and $Y = 0$ otherwise. Then,

$$\mathbb{E}[N^2|Y = 1] = 1 \qquad \text{(because } N = 1 \text{ if } Y = 1\text{)}$$

$$\mathbb{E}[N^2|Y = 0] = \mathbb{E}[(1 + N)^2],$$

Examples

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

and thus

$$\begin{aligned}\mathbb{E}[N^2] &= \mathbb{E}[N^2|Y=1]\,\mathbb{P}\{Y=1\} + \mathbb{E}[N^2|Y=0]\,\mathbb{P}\{Y=0\} \\ &= 1\cdot p + (1+\mathbb{E}[2N+N^2])(1-p) \\ &= p + \frac{2(1-p)}{p} + (1-p)\,\mathbb{E}[N^2], \qquad \text{(because } \mathbb{E}[N]=1/p)\end{aligned}$$

therefore,

$$\mathbb{E}[N^2] = \frac{2-p}{p^2}, \quad \mathrm{Var}(N) = \mathbb{E}[N^2] - (\mathbb{E}[N])^2 = \frac{1-p}{p^2}. \qquad \blacksquare$$

Computing probabilities by conditioning

- Suppose that we want to calculate $\mathbb{P}(E)$, which is not easy to compute.
- However, we find another random variable Y, such that $\mathbb{P}(E|Y = y)$ is known.

Proposition 35

We have

$$\mathbb{P}(E) = \begin{cases} \sum_y \mathbb{P}(E|Y = y) p_Y(y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{P}(E|Y = y) f_Y(y) dy & \text{if } Y \text{ is continuous} \end{cases}$$

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Example 36 (The best-prize problem)

Suppose that we are to be presented with n distinct prizes, in sequence. After being presented with a prize, we must immediately decide whether to accept it or to reject it and consider the next prize. The only information we are given when deciding whether to accept a prize is the relative rank of that prize compared to ones already seen. That is, for instance, when the fifth prize is presented, we learn how it compares with the four prizes we've already seen. Suppose that once a prize is rejected, it is lost, and that our objective is to maximize the probability of obtaining the best prize. Assuming that all $n!$ orderings of the prizes are equally likely, how well can we do?

Solution.

Fix a value k, where $0 \leqslant k < n$, and consider the strategy that rejects the first k prizes and then accepts the first one that is better than all of those first k. Let $\mathbb{P}_k(A)$ denote the probability that the best prize is selected when this strategy is employed. To compute the probability, conditional on X (the position of the best prize),

$$\mathbb{P}_k(A) = \sum_{i=1}^{n} \mathbb{P}_k(A|X = k)\, \mathbb{P}\{X = i\} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_k(A|X = i).$$

Now, if $i \leqslant k$, then $\mathbb{P}_k(A|X = i) = 0$. On the other hand, if $i > k$,

$$\mathbb{P}_k(A|X = i) = \mathbb{P}\{\text{best of first } i - 1 \text{ is among the first } k|X = i\} = \frac{k}{i - 1}.$$

Solution (Cont'd).

From the preceding, we obtain

$$\mathbb{P}_k(A) = \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1} \approx \frac{k}{n} \log\left(\frac{n}{k}\right).$$

To maximize the probability, we choose $k = n/e$, and it follows that, in this case,

$$\mathbb{P}_k(A) \approx \frac{1}{e} \approx 0.36788.$$

■

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 37

Let U be a uniform random variable on $(0, 1)$, and suppose that the conditional distribution of X, given that $U = p$, is binomial with parameters n and p. Find the probability mass function of X.

Examples

Solution.

Conditioning on U gives

$$\begin{aligned}
\mathbb{P}\{X = i\} &= \int_0^1 \mathbb{P}\{X = i | U = u\} f_U(u) du \\
&= \int_0^1 \mathbb{P}\{X = i | U = u\} du \\
&= \frac{n!}{i!(n-i)!} \int_0^1 u^i (1-u)^{n-i} du = \frac{n!}{i!(n-i)!} \frac{\Gamma(i+1)\Gamma(n-i+1)}{\Gamma(n+2)} \\
&= \frac{1}{n+1}, \quad i = 0, \ldots, n.
\end{aligned}$$

■

Example 38

Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute $\mathbb{P}\{X < Y\}$.

Solution.

Conditioning on Y yields

$$\begin{aligned}\mathbb{P}\{X < Y\} &= \int_{-\infty}^{\infty} \mathbb{P}\{X < Y | Y = y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \mathbb{P}\{X < y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y} f_X(x) f_Y(y) dx dy.\end{aligned}$$

■

Conditional variance

Definition 39

The conditional variance of X given that $Y = y$ is defined as

$$\mathrm{Var}(X|Y) = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y].$$

Remark

We have

$$\mathrm{Var}(X|Y) = \mathbb{E}[X^2|Y] - (\mathbb{E}[X|Y])^2.$$

Total variance formula

Proposition 40

We have

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y]).$$

Example 41 (Sum of a random number of random variables)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2, and let N be a nonnegative integer-valued random variable that is independent of all others. Find $\mathrm{Var}(\sum_{i=1}^{N} X_i)$.

Solution.

Condition on N:

$$\mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right] = N\,\mathbb{E}[X] = N\mu,$$

$$\operatorname{Var}\left(\sum_{i=1}^{N} X_i \,\middle|\, N\right) = N\operatorname{Var}(X) = N\sigma^2.$$

Then,

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{N} X_i\right) &= \mathbb{E}\left[\operatorname{Var}\left(\sum_{i=1}^{N} X_i \,\middle|\, N\right)\right] + \operatorname{Var}\left(\mathbb{E}\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right)\\ &= \mathbb{E}[\sigma^2 N] + \operatorname{Var}(\mu N) = \sigma^2\,\mathbb{E}[N] + \mu^2\operatorname{Var}(N).\end{aligned}$$

■

Moment generating function

Defintion

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Definition 42

The moment generating function $M : \mathbb{R} \to \mathbb{R}$ of the random variable X is defined as

$$M(t) = \mathbb{E}[e^{tX}] \quad \text{for all } t \in \mathbb{R}.$$

Remark (Why it is called moment generating function?)

Note that

$$\begin{aligned} M'(t) &= \frac{d}{dt}\mathbb{E}[e^{tX}] \\ &= \mathbb{E}\left[\frac{d}{dt}(e^{tX})\right] \qquad \text{(We assume that } \mathbb{E} \text{ and } d/dt \text{ can be interchanged)} \\ &= \mathbb{E}[Xe^{tX}]. \end{aligned}$$

Then, $M'(0) = \mathbb{E}[X]$. Similarly, $M''(0) = \mathbb{E}[X^2]$, and so on.

Proposition 43

If $X \sim \text{Binomial}(n, p)$, then

$$M(t) = (pe^t + 1 - p)^n.$$

Proof.

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n. \end{aligned}$$

∎

Exponential distribution

Proposition 44

If $X \sim \text{Exp}(\lambda)$, then

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tX}] \\ &= \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx \\ &= \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda. \end{aligned}$$

We note from this derivation that, for the exponential distribution, $M(t)$ is defined only for $t < \lambda$.

Proposition 45

If $X \sim N(\mu, \sigma^2)$, then

$$M(t) = e^{\mu t - \frac{\sigma^2 t^2}{2}}.$$

Properties of moment generating functions

Proposition 46

If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

One-to-one

Proposition 47

The moment generating function uniquely determines the distribution.

Example 48

If $M_X(t) = (1/2)^{10}(e^t + 1)^{10}$, then

$$X \sim \text{Binomial}(10, \frac{1}{2}).$$

If $M_Y(t) = \frac{1}{1-t}$ for $t < 1$, then $Y \sim \text{Exp}(1)$.

Joint moment generating function

Definition 49

For any two random variables X and Y, the joint moment generating function $M(s,t)$ of (X,Y) is defined as

$$M(s,t) = \mathbb{E}[e^{sX+tY}].$$

Proposition 50

We have

$$M_X(s) = M(s,0).$$

Random variables X and Y are independent, if and only if

$$M(s,t) = M_X(s)M_Y(t).$$

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY 南方科技大学

Example 51

If X and Y are i.i.d. from $N(\mu, \sigma^2)$. Find the joint distribution of $X+Y$ and $X-Y$.

Solution.

Let M be the moment generating function of $N(\mu, \sigma^2)$. Note that

$$
\begin{aligned}
\mathbb{E}[e^{s(X+Y)+t(X-Y)}] &= \mathbb{E}[e^{(s+t)X+(s-t)Y}] \\
&= M(s+t)M(s-t) \\
&= e^{2\mu s+\sigma^2 s^2} e^{\sigma^2 t^2}
\end{aligned}
$$

which is the product of the moment generating functions of $N(2\mu, 2\sigma^2)$ and $N(0, 2\sigma^2)$. Therefore, $X+Y \sim N(2\mu, 2\sigma^2)$ and $X-Y \sim N(0, 2\sigma^2)$ are independent. ■

Further reading

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

[1] Sheldon M. Ross (谢尔登・M. 罗斯).

A first course in probability (概率论基础教程): Chapter 7.

10th edition (原书第十版), 机械工业出版社

[2] Sheldon M. Ross (谢尔登・M. 罗斯).

Introduction to Probability Models (概率模型导论): Chapter 3.

12th edition (原书第十二版), 人民邮电出版社