Expectation

Math & Stat for Data Science

Graduate School of Data Science
Seoul National University

Expectation

- Suppose researchers are interested in the height of Korean population
  - How can we summarize the data?

Expectation

- Now researchers are interested in both height and weight
  - How we can capture the relationship between these two?

Expectation

3.1 Definition. *The* **expected value**, *or* **mean**, *or* **first moment**, *of* X *is defined to be*

$$\mathbb{E}(X) = \int x\, dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases} \tag{3.1}$$

assuming that the sum (or integral) is well defined. We use the following notation to denote the expected value of X*:*

$$\mathbb{E}(X) = \mathbb{E}X = \int x\, dF(x) = \mu = \mu_X . \tag{3.2}$$

- One-number summary of the distribution
- Can be approximated by sample mean (for IID samples)

Expectation (Binary Dist)

- X~ Bernoulli(p)
  - E(X) = p

Expectation (Binary Dist)

- X~Binomial(n,p)
  - E(X) = np

Expectation (Binary Dist)

- X~Poisson(λ)
  - E(X) = λ

Expectation

3.6 Theorem (The Rule of the Lazy Statistician). *Let* $Y = r(X)$. *Then*

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x) dF_X(x). \tag{3.3}$$

- Expected value of transformed variables can be easily calculated using the above theorem.
- For binary variables (or multivariate variables)

$$\mathbb{E}(r(X, Y)) = \int \int r(x, y) dF(x, y)$$

Example

3.7 Example. Let $X \sim \text{Unif}(0,1)$. Let $Y = r(X) = e^X$.

E(r(X))?

Example

3.9 Example. Let (X, Y) have a jointly uniform distribution on the unit square. Let $Z = r(X, Y) = X^2 + Y^2$. Then,

E(r(X, Y))?

Properties of Expectation

3.11 Theorem. *If $X_1, \ldots, X_n$ are random variables and $a_1, \ldots, a_n$ are constants, then*

$$\mathbb{E}\left(\sum_i a_i X_i\right) = \sum_i a_i \mathbb{E}(X_i). \tag{3.5}$$

- Very useful property!
- Do not require independence of X!!
- Example: mean of Binomial (n, p)?

Properties of Expectation

3.13 Theorem. *Let $X_1, \ldots, X_n$ be independent random variables. Then,*

$$\mathbb{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_i \mathbb{E}(X_i). \tag{3.6}$$

- Requires independence of X
- Example: Mean of XY, where X and Y are independent and X~Bernoulli(p_1) and Y~ Bernoulli(P_2)

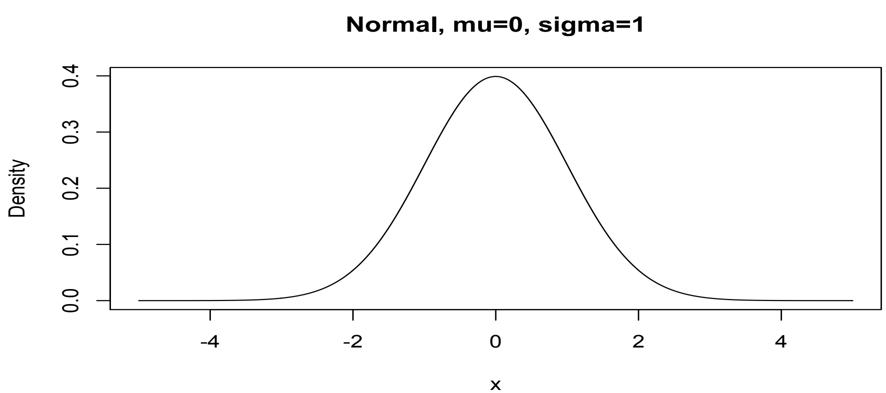
Variance and Covariance

3.14 Definition. *Let X be a random variable with mean μ. The* **variance** *of X — denoted by σ^2 or σ_X^2 or $\mathbb{V}(X)$ or $\mathbb{V}X$ — is defined by*

$$\sigma^2 = \mathbb{E}(X-\mu)^2 = \int (x-\mu)^2 dF(x) \tag{3.7}$$

assuming this expectation exists. The **standard deviation** *is* $\mathrm{sd}(X) = \sqrt{\mathbb{V}(X)}$ *and is also denoted by σ and σ_X.*

- Variance represents the spread of distribution



Variance – Important properties

3.15 Theorem. *Assuming the variance is well defined, it has the following properties:*

1. $\mathbb{V}(X) = \mathbb{E}(X^2) - \mu^2$.

2. *If a and b are constants then* $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.

3. *If $X_1, \ldots, X_n$ are independent and $a_1, \ldots, a_n$ are constants, then*

$$\mathbb{V}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathbb{V}(X_i). \tag{3.8}$$

Variance

- Example: Variance of binomial(n, p)?

Variance

- Sample mean and variance

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

3.17 Theorem. *Let $X_1, \ldots, X_n$ be* IID *and let $\mu = \mathbb{E}(X_i)$, $\sigma^2 = \mathbb{V}(X_i)$. Then*

$$\mathbb{E}(\overline{X}_n) = \mu, \quad \mathbb{V}(\overline{X}_n) = \frac{\sigma^2}{n} \quad \text{and} \quad \mathbb{E}(S_n^2) = \sigma^2.$$

Covariance and Correlation

3.18 Definition. *Let X and Y be random variables with means μ_X and μ_Y and standard deviations σ_X and σ_Y. Define the* **covariance** *between X and Y by*

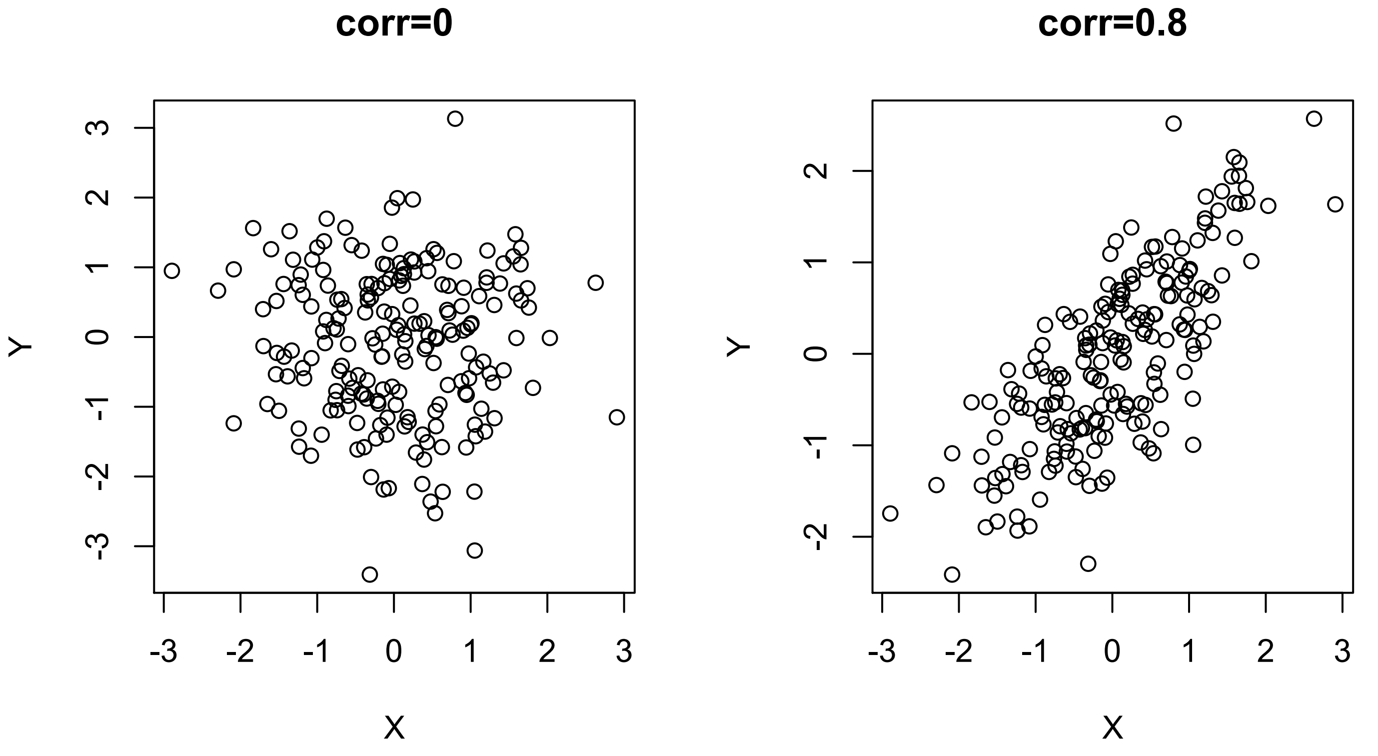
$$\mathsf{Cov}(X,Y) = \mathbb{E}\Big((X-\mu_X)(Y-\mu_Y)\Big) \tag{3.11}$$

and the **correlation** *by*

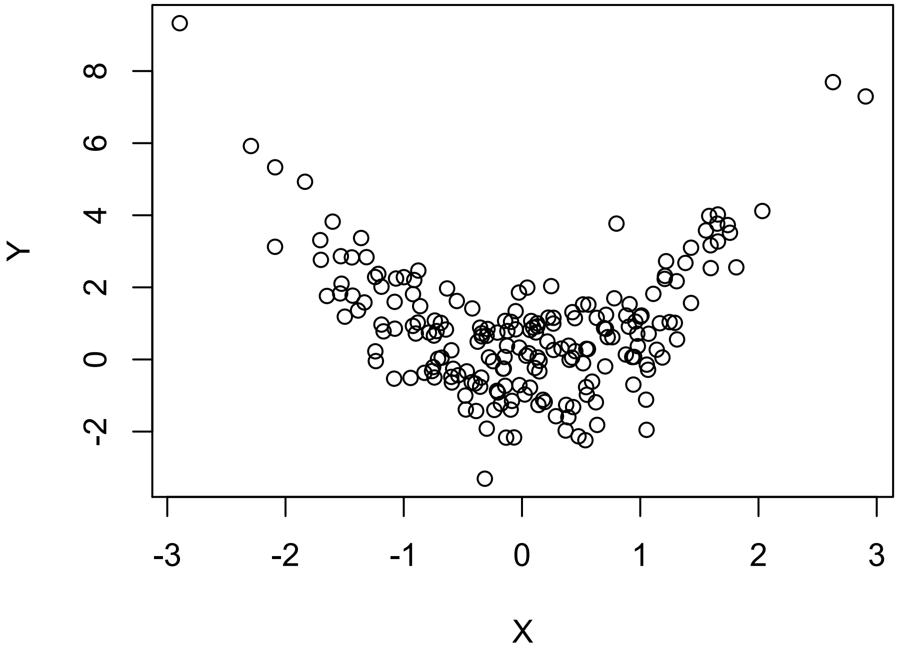
$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sigma_X \sigma_Y}. \tag{3.12}$$

- Indicate the strength of linear relationship between two random variables X and Y

Covariance and Correlation



Covariance and Correlation



Correlation?

Covariance and Correlation

3.19 Theorem. *The covariance satisfies:*

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

The correlation satisfies:

$$-1 \leq \rho(X,Y) \leq 1.$$

Covariance and Correlation

3.20 Theorem. $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathsf{Cov}(X,Y)$ *and* $\mathbb{V}(X-Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\mathsf{Cov}(X,Y)$. *More generally, for random variables* $X_1, \ldots, X_n$,

$$\mathbb{V}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathbb{V}(X_i) + 2 \sum\sum_{i<j} a_i a_j \mathsf{Cov}(X_i, X_j).$$

Important RVs

Distribution	Mean	Variance
Point mass at a	a	0
Bernoulli(p)	p	$p(1-p)$
Binomial(n,p)	np	$np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ
Uniform(a,b)	$(a+b)/2$	$(b-a)^2/12$
Normal(μ,σ^2)	μ	σ^2
Exponential(β)	β	β^2
Gamma(α,β)	$\alpha\beta$	$\alpha\beta^2$
Beta(α,β)	$\alpha/(\alpha+\beta)$	$\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$
t_ν	0 (if $\nu > 1$)	$\nu/(\nu-2)$ (if $\nu > 2$)
χ^2_p	p	$2p$
Multinomial(n,p)	np	see below
Multivariate Normal(μ,Σ)	μ	Σ

Multivariate & Conditional Expectation & Monte-Carlo approach

Multivariate RVs

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_k) \end{pmatrix}$$

Mean Vector

$$\mathbb{V}(X) = \begin{bmatrix} \mathbb{V}(X_1) & \mathsf{Cov}(X_1, X_2) & \cdots & \mathsf{Cov}(X_1, X_k) \\ \mathsf{Cov}(X_2, X_1) & \mathbb{V}(X_2) & \cdots & \mathsf{Cov}(X_2, X_k) \\ \vdots & \vdots & \vdots & \vdots \\ \mathsf{Cov}(X_k, X_1) & \mathsf{Cov}(X_k, X_2) & \cdots & \mathbb{V}(X_k) \end{bmatrix}.$$

Covariance Matrix

Multivariate RVs

3.21 Lemma. *If a is a vector and X is a random vector with mean μ and variance Σ, then $\mathbb{E}(a^T X) = a^T \mu$ and $\mathbb{V}(a^T X) = a^T \Sigma a$. If A is a matrix then $\mathbb{E}(AX) = A\mu$ and $\mathbb{V}(AX) = A\Sigma A^T$.*

- Very useful to identify the mean and variance of linear combination of random variables

Multivariate RVs

- Example: Suppose $(X_1, X_2, X_3)^T$ have the following mean and variance

$$\mu = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix},$$

  - Let $Y = X_1 + 0.5X_2 + 0.5X_3$. Mean and variance of Y?

Conditional Expectation

3.22 Definition. *The conditional expectation of X given $Y = y$ is*

$$\mathbb{E}(X|Y=y) = \begin{cases} \sum x\, f_{X|Y}(x|y)\, dx & \text{discrete case} \\ \int x\, f_{X|Y}(x|y)\, dx & \text{continuous case.} \end{cases} \tag{3.13}$$

If $r(x,y)$ is a function of x and y then

$$\mathbb{E}(r(X,Y)|Y=y) = \begin{cases} \sum r(x,y)\, f_{X|Y}(x|y)\, dx & \text{discrete case} \\ \int r(x,y)\, f_{X|Y}(x|y)\, dx & \text{continuous case.} \end{cases} \tag{3.14}$$

Conditional Expectation

- Given Y, what is the expected values of X?
  - Ex. Given Height=180 cm, what is the expected value of Weight

- Important: $E(X|Y=y)$ & $E(r(X,Y)|Y=y)$ are functions of y. So $E(X|Y)$ is a random variable of Y
  - Ex. $X \sim \text{Uniform}(0,1)$, and $Y \sim \text{Uniform}(x,1)$ given $X=x$. Then $E(Y \mid X) = (1+x)/2$

Conditional Expectation

3.24 Theorem (The Rule of Iterated Expectations). *For random variables X and Y, assuming the expectations exist, we have that*

$$\mathbb{E}\left[\mathbb{E}(Y|X)\right] = \mathbb{E}(Y) \quad \text{and} \quad \mathbb{E}\left[\mathbb{E}(X|Y)\right] = \mathbb{E}(X). \tag{3.15}$$

More generally, for any function $r(x, y)$ we have

$$\mathbb{E}\left[\mathbb{E}(r(X,Y)|X)\right] = \mathbb{E}(r(X,Y)). \tag{3.16}$$

Conditional Variance

3.26 Definition. *The* **conditional variance** *is defined as*

$$\mathbb{V}(Y|X = x) = \int (y - \mu(x))^2 f(y|x)dy \tag{3.17}$$

where $\mu(x) = \mathbb{E}(Y|X = x)$.

Conditional Variance is also a RV

3.27 Theorem. *For random variables* X *and* Y,

$$\mathbb{V}(Y) = \mathbb{E}\mathbb{V}(Y|X) + \mathbb{V}\mathbb{E}(Y|X).$$

Conditional Expectation and Variance

- Example: Suppose height of male ~ $N(173, 5^2)$ and female ~ $N(163, 4^2)$. The numbers of males and females are the same
  - Mean height?

Conditional Expectation and Variance

- Example: Suppose height of male ~ $N(173, 5^2)$ and female ~ $N(163, 4^2)$. The numbers of males and females are the same
  - Variance?

Monte-Carlo simulation

- Suppose we want to calculate E(f(x)) and V(f(x)), where distribution of x is know and easily sampled
  - EX. x ~ N(0,1) and f(x) = x^3

- In many situations, it is difficult to get them analytically

- Monte Carlo approach can be used
  - Simulate x, B times, $x_1, \dots, x_B$
  - Estimate E(f(x)) as the sample mean

$$E\big(f(x)\big) \approx \frac{1}{n}\sum_{i=1}^{B} f(x_i)$$

Monte-Carlo simulation

- EX. Suppose that (height,weight) in Korean Male follows MVN with

$$\mu = \binom{173}{68}, \Sigma = \begin{pmatrix} 5^2 & 10 \\ 10 & 4^2 \end{pmatrix},$$

  - Mean and variance of BMI (kg/m^2)?

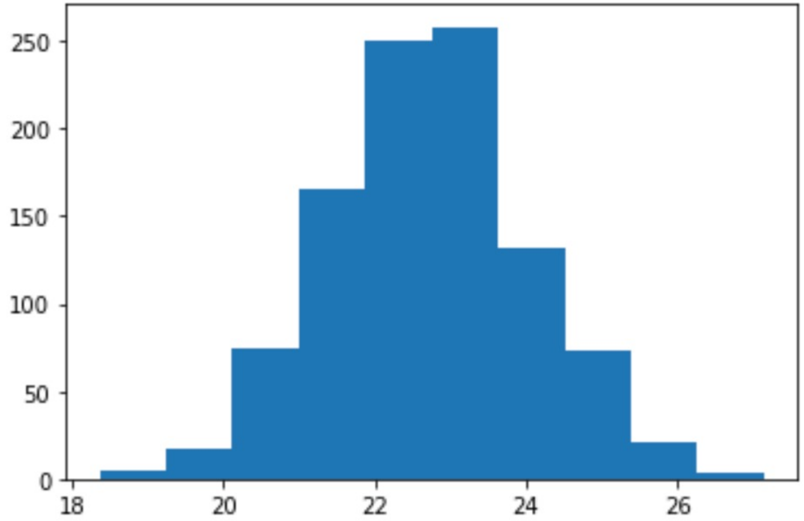
Monte-Carlo simulation: BMI

```
import numpy as np
import matplotlib.pyplot as plt


mean = np.array([173, 68])
cov = np.array([[25,10], [10,16]])

X = np.random.multivariate_normal(mean, cov, size=1000)
BMI = X[:,1]/(X[:,0]/100) **2

plt.hist(BMI)
```




```
print('mean:', np.mean(BMI))
print('sd:', np.std(BMI))
```

```
mean: 22.707407367106246
sd: 1.3214436859878371
```

MGF

Moment Generating Function

3.29 Definition. *The* moment generating function MGF, *or* Laplace transform, *of X is defined by*

$$\psi_X(t) = \mathbb{E}(e^{tX}) = \int e^{tx} dF(x)$$

where t varies over the real numbers.

- Have all the information on the distribution
- Useful to derive the distribution of sum of random variables
- Moment calculation: $\psi^{(k)}(0) = \mathbb{E}(X^k)$

Moment Generating Function

3.31 Lemma. *Properties of the* MGF.

(1) If $Y = aX + b$, *then* $\psi_Y(t) = e^{bt}\psi_X(at)$.

(2) If $X_1, \ldots, X_n$ *are independent and* $Y = \sum_i X_i$, *then* $\psi_Y(t) = \prod_i \psi_i(t)$ *where* ψ_i *is the* MGF *of* X_i.

MGF

- Example:
    - Let $X \sim \text{Binomial}(n, p)$. MGF of X?

MGF

- Example:
  - Let X ~ Binomial(n_1, p), and Y~ Binomial (n_2, p). MGF of X+Y?

MGF

3.33 Theorem. *Let X and Y be random variables. If $\psi_X(t) = \psi_Y(t)$ for all t in an open interval around 0, then $X \stackrel{d}{=} Y$.*

- MGF is widely used when to derive the distribution!

MGF

Moment Generating Functions for Some Common Distributions

Distribution	MGF $\psi(t)$
Bernoulli(p)	$pe^t + (1-p)$
Binomial(n, p)	$(pe^t + (1-p))^n$
Poisson(λ)	$e^{\lambda(e^t-1)}$
Normal(μ,σ)	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$
Gamma(α,β)	$\left(\frac{1}{1-\beta t}\right)^\alpha$ for $t < 1/\beta$

MGF

- There are slightly different versions of similar functions
  - Cumulant Generating Function
  - Characteristic function

MGF

- Using MGF (CGF and characteristic functions), distribution function can be estimated
  - This kind of technique can be very useful…

Summary

- Expectation
  - One-number summary of the distribution
  - Linear operator
- Variance
  - Represent the spread of distribution
- Covariance & Correlation
  - Indicate the (linear) relationship between two RV

- Conditional Expectation
- Moment generating function