chapter five

SUM OF RANDOM VARIABLES AND LONG-TERM AVERAGES

Let $X_1, X_2, \ldots, X_n$ be the sequence of random variables, and let S_n be their sum:

$$S_n = X_1 + X_2 + \cdots + X_n$$

The expected value:

Regardless of statistical dependence, the expected value of a sum of n random variables is equal to the sum of the expected values:

$$E[S_n] = E[X_1 + X_2 + \cdots + X_n]$$

$$= E[X_1] + E[X_2] + \cdots + E[X_n]$$

$$= \mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n}.$$

The Variance:

$$VAR[S_n] = VAR[X_1 + X_2 + \cdots + X_n]$$

$$= E\left[\sum_{j=1}^{n} (X_j - E[X_j]) \sum_{k=1}^{n} (X_k - E[X_k])\right]$$

$$\sigma^2_{X_1+X_2+\cdots+X_n} = \sum_{j=1}^{n}\sum_{k=1}^{n} E\Big[\big(X_j - E[X_j]\big)\big(X_k - E[X_k]\big)\Big]$$

$$= \sum_{k=1}^{n} VAR[X_k] + \sum_{\substack{j=1 \\ j\neq k}}^{n}\sum_{k=1}^{n} COV(X_j, X_k)$$

$$= \underbrace{\sum_{k=1}^{n} \sigma^2_{X_k}}_{\text{individual variances}} + \sum_{\substack{j=1 \\ j\neq k}}^{n}\sum_{k=1}^{n} \underbrace{\sigma_{X_j X_k}}_{\text{Covariances}}$$

Therefore, in general, the variance of the sum of random variables is not equal to the sum of the individual variances.

Special Case:

When X_i's are independent random variables, Then,

$$COV[X_j, X_k] = \sigma_{X_j X_k} = 0 \quad , \text{for } j \neq k,$$

and

$$\sigma^2_{X_1+X_2+\cdots+X_n} = \sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_n}.$$

Example:

Find the mean and the variance of the sum of n independent and identically distributed (iid) random variables each with mean μ and variance σ^2:

$$E[S_n] = E[X_1 + X_2 + \cdots + X_n]$$
$$= \mu + \mu + \cdots + \mu$$
$$= n\mu$$

and

$$VAR[S_n] = \sigma_{S_n}^2$$
$$= n\sigma_{X_i}^2 = n\sigma^2.$$

PDF of Sums of Independent Random Variables

We can show that the transform methods can be used to find the pdf of $S_n = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent.

Let us consider the $n=2$ case, $Z = X + Y$, X and Y are independent random variables. The characteristic function of Z is given by

$$\Phi_Z(\omega) = E\left[e^{j\omega Z}\right]$$

$$= E\left[e^{j\omega(X+Y)}\right]$$

$$= E\left[e^{jwX}\, e^{jwY}\right]$$

• Since X and Y are independent random variables.

$$= \underbrace{E\left[e^{jwX}\right]}_{\Phi_X(\omega)}\underbrace{E\left[e^{jwY}\right]}_{\Phi_Y(\omega)}$$

$\Phi_Z(\omega)$ can be viewed as the Fourier transform of the pdf of the random variable Z,

$$\Phi_Z(\omega) = \mathcal{F}\left[f_Z(z)\right].$$

$f_Z(z)$ is given by the convolution of the pdf of X and Y

$$f_Z(z) = f_X(x) \underset{\text{(Convolution)}}{*} f_Y(y)$$

The Sample Mean and The Laws of Large Numbers

Let X be a random variable for which the mean, $E[E] = \mu$, is unknown.

Let $X_1, X_2, \dots, X_2$ denote n idependent repeated measurements of X, that is, the X_j's are independent, identically distributed (iid) random variables with the same pdf of X.

Sample Mean:

The sample mean of the squence is used to estimate $E[X]$:

$$M_n = \frac{1}{n} \sum_{j=1}^{n} X_j \; .$$

Theorem: Weak-Law of Large Numbers (Khinchin's Theorem, 1929)

Let $X_1, X_2, \dots$ be a sequence of iid random variables with mean, $E[X]$, then for $\epsilon > 0$,

$$\lim_{n\to\infty} P_r \left\{ \left| M_n - \mu \right| < \epsilon \right\} = 1 \; . \qquad (A)$$

The weak-law of large numbers states that for a large enough fixed value of n, the sample mean using n samples will be close to the true mean with a high probability.

Do we have an answer this question?

The weak-law of large numbers does not adress the question about what happens to the sample mean as a function of n if we make additional measurements.

This question is taken up by the strong law of large numbers, which we discuss next:

Theorem: STRONG LAW OF LARGE NUMBERS

- Suppose we make a series of independent measurments of the same random variable.
- Let $X_1, X_2, \ldots$ be the resulting sequence of iid random variables with mean μ.

Consider the sequence of sample means that results from the above measurements:

$$M_1 = X_1$$

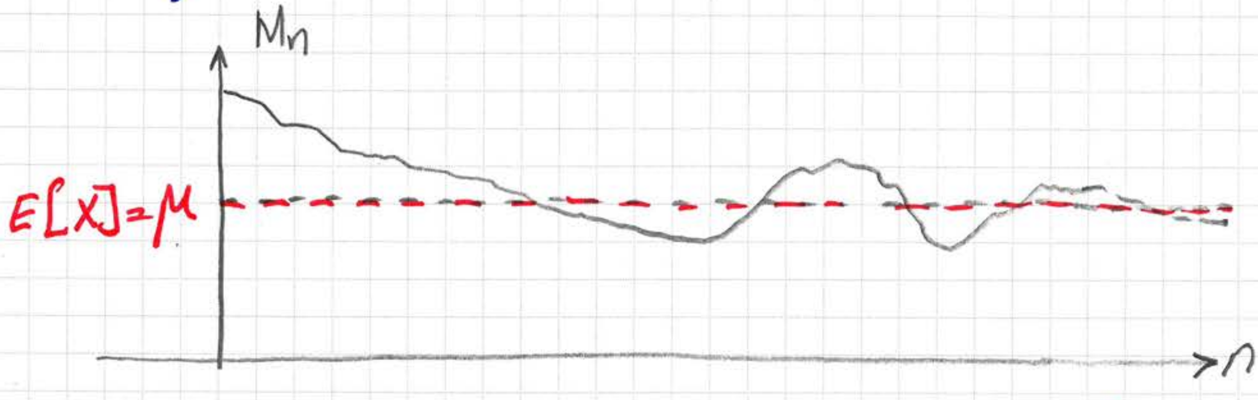
$$M_2 = \frac{1}{2}(X_1 + X_2)$$

$$M_3 = \frac{1}{3}(X_1 + X_2 + X_3)$$

$$\vdots$$

$$M_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

$$\vdots$$



- Each particular sequence of sample means $M_1, M_2, M_3, \ldots$ converges to μ, that is, we expect that with a high probability,
  - each particular sequence of sample means approaches to μ and stays there.

The statistical regularity notion that we discussed in Chapter 1 leads to this expectation or convergence.

Strong Law of Large Numbers:

Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite mean $E[X]=\mu<\infty$ and finite Variance $\sigma_X^2<\infty$, then

$$\Pr\{\lim_{n\to\infty} M_n = \mu\} = 1. \tag{B}$$

. This equation appears similar to equation

$$\lim_{n\to\infty} \Pr\{|M_n-\mu|<\epsilon\} = 1, \text{ for } \epsilon>0.$$

but it makes a dramatically different statements.

This new equation (B) states that with probability 1, every sequence of sample mean calculations will eventually approach and stay close to $E[X]=\mu$.

In physical situations where statistical regularity holds, we expect this type of convergence.

The Central Limit Theorem

Let

$$S_n = X_1 + X_2 + \dots + X_n$$

be the sum n iid random variables with finite mean $E[X] = \mu < \infty$ and finite variance, $\sigma^2 < \infty$, and let Z_n be zero-mean and unit variance random variable defined by

$$Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}}$$

then

$$\lim_{n\to\infty} Pr\{Z_n \leq z\} = \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx}_{\text{This is the cdf of the standard normal (Gaussian) random variable.}} .$$

Proof: It is given in the texbook and will not be repeated here.

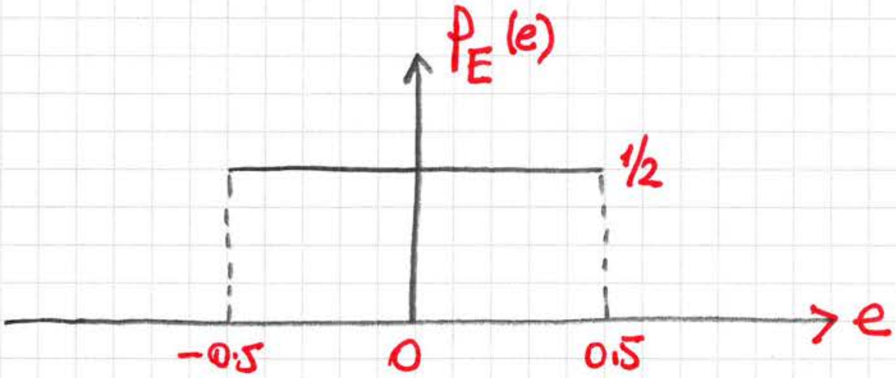
This is the cdf of the standart normal (Gaussian) random variable.

From this result, we understand that n becomes large, the cdf of a properly normalized S_n, namely Z_n, approaches a Baussian random variable.

There are several examples in the book to verify this theorem.

Example 1:

We are summing 100 numbers by rounding the closest integer. Each rounding error is independent and uniformly distributed in the interval $(-0.5, 0.5)$,



$$\underbrace{X_i}_{\text{The original ith number}} = \underbrace{X_{q_i}}_{\text{The rounded to closest integer}} + \underbrace{E_i}_{\text{Rounding error}}$$

The sum of errors

$$S_{100} = E_1 + E_2 + \cdots + E_{100}$$

and

$$E_i \sim \mathrm{Unif}(-0.5, 0.5), \quad i = 1, 2, \ldots, 100.$$

According to the central limit theorem, the random variable S_{100} is gaussian (normal) distributed.

$$\mu_{S_{100}} = 100\, \mu_{E_i}$$

and

$$\sigma^2_{S_{100}} = 100\, \sigma^2_{E_i} \quad .$$

From $f_E(e)$, we can calculate

$$\mu_{E_i} = \int_{-0.5}^{0.5} e\, f_E(e)\, de = \int_{-0.5}^{0.5} e\, \frac{1}{2}\, de = 0$$

and

$$\sigma^2_{E_i} = \int_{-0.5}^{0.5} (e-0)^2 f_E(e)\, de$$

$$= \int_{-0.5}^{0.5} e^2\, \frac{1}{2}\, de = \frac{1}{12} \quad .$$

Then,

$$\mu_{S_{100}} = 100\, \mu_{E_i}$$

$$= 100 \cdot (0) = 0$$

and

$$\sigma^2_{S_{100}} = 100\, \sigma^2_{E_i}$$

$$= 100 \cdot \left(\frac{1}{12}\right) = \frac{100}{12} \; .$$

Question ?

$$Pr\{|S_{100}| > 5\} = ?$$

The absolute value of the sum of the errors.

$$Pr\{|S_{100}|>5\} = 1 - Pr\{|S_{100}| \leq 5\}$$

$$= 1 - Pr\{-5 \leq S_{100} \leq 5\}$$

$$= 1 - \left[\Phi\left(\frac{5-0}{\sqrt{\frac{100}{12}}}\right) - \Phi\left(\frac{-5-0}{\sqrt{\frac{100}{12}}}\right)\right]$$

$$= 1 - \left[\Phi(1.73) - \Phi(-1.73)\right]$$

$$= 1 - [0.9582 - 0.0418] = 0.08.$$

Example:

Within an hour, the average number of costumers coming into a bank branch is 30. What is the probability of coming more then 35 costumbers?

X random variable shows the number of costumers visiting a bank within a hour is a Poisson distribution, $X \sim Poisson(30)$.

$$Pr\{X > 35\} = P_{36} + P_{37} + P_{38} + \cdots$$

$$= 1 - \sum_{k=0}^{35} P_i$$

$$= 1 - \sum_{k=0}^{35} \frac{30^k}{k!} e^{-30} .$$

Instead of calculating this expression, we can use the normal distribution approximation, (Remember $E[X] = \sigma_X^2 = 30$),

$$Pr\{X > 35\} = 1 - \Phi\left(\frac{35-30}{\sqrt{30}}\right)$$

$$= 1 - \Phi(0.91)$$

$$= 1 - 0.8186$$

$$= 0.1814 .$$

Example [18]:

Suppose that a random variable X has a continuous uniform distribution

$$f_X(x) = \begin{cases} 1/2 & , \ 4 \le x \le 6 \\ 0 & , \ \text{otherwise} . \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$, namely,

$$M_{40} = \frac{1}{40}\left(X_1 + X_2 + \cdots + X_{40}\right).$$

The mean and variance of X are $\mu = 5$ and $\sigma^2 = (6-4)^2/12 = 1/3$. The central limit theorem indicates that the distribution of M_{40} is approximately normal with mean

$$\mu_{M_{40}} = \frac{1}{40}\left(\mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_{40}}\right) = \frac{1}{40}(5+5+\cdots+5) = 5$$

and variance

$$\sigma_{M_{40}}^2 = \frac{1}{40^2}\left(\sigma_{X_1}^2 + \sigma_{X_2}^2 + \cdots + \sigma_{X_{40}}^2\right) = \frac{1}{40}\left(\frac{1}{3} + \frac{1}{3} + \cdots \frac{1}{3}\right) = \frac{1/3}{1/40} = \frac{1}{120}$$

and $M_{40} \sim \mathcal{N}\left(5, \frac{1}{120}\right)$