Henry Stark and John W. Woods,
Probability, Statistics, and Random Variables for Engineers, 4th ed.,
Pearson Education Inc., 2012. ISBN: 978-0-13-231123-6

Chapter 4 Expectation and Moments

Sections

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

4.1 Expected Value of a Random Variable 215

Mean Value: the expected mean value of measurements of a process involving a random variable.

This is commonly called the expectation operator or expected value of …
and is mathematically described as:

$$\overline{X} = E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx$$

For laboratory experiments, the expected value of a voltage measurement can be thought of as the DC voltage.

For discrete random variables, the integration becomes a summation and

$$\overline{X} = E[X] = \sum_{x=-\infty}^{\infty} x \cdot f_X(x) = \sum_{x=-\infty}^{\infty} x \cdot \Pr(X = x)$$

General concept of an expected value

In general, the expected value of a function is:

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) \cdot f_X(x) \cdot dx$$

$$E[g(X)] = \sum_{x=-\infty}^{\infty} g(X) \cdot f_X(x) = \sum_{x=-\infty}^{\infty} g(X) \cdot \Pr(X = x)$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) \cdot f_X(x) \cdot dx$$

Estimating a parameter:

If we know the expected value, you have a simple *estimate* of future expected outcomes.

$$\hat{x} = \overline{X} = E[X]$$

Or for $y = g(x)$

$$\hat{y} = E[y] = E[g(X)]$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Example 4.1-2 Expected value of a Gaussian

$$\overline{X} = E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx$$

$$f_X(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-(x-\mu)^2}{2 \cdot \sigma^2}\right), \textit{ for } -\infty < x < \infty$$

$$\overline{X} = E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-(x-\mu)^2}{2 \cdot \sigma^2}\right) \cdot dx$$

Letting $z = \frac{x-\mu}{\sigma}$ with $dz = \frac{dx}{\sigma}$

$$\overline{X} = E[X] = \int_{-\infty}^{\infty} (z \cdot \sigma + \mu) \cdot \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-z^2}{2}\right) \cdot \sigma \cdot dz$$

$$\overline{X} = E[X] = \mu \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-z^2}{2}\right) \cdot dz + \int_{-\infty}^{\infty} \frac{z \cdot \sigma}{\sqrt{2\pi \cdot}} \cdot \exp\left(\frac{-z^2}{2}\right) \cdot dz$$

$$\overline{X} = E[X] = \mu + \frac{\sigma}{\sqrt{2\pi \cdot}} \cdot \int_{-\infty}^{\infty} z \cdot \exp\left(\frac{-z^2}{2}\right) \cdot dz$$

$$\overline{X} = E[X] = \mu - \frac{\sigma}{\sqrt{2\pi \cdot}} \cdot \exp\left(\frac{-z^2}{2}\right)\Bigg|_{-\infty}^{\infty}$$

$$\overline{X} = E[X] = \mu$$

Moments

The moments of a random variable are defined as the expected value of the powers of the measured output or …

$$\overline{X^n} = E\left[X^n\right] = \int_{-\infty}^{\infty} x^n \cdot f_X(x) \cdot dx$$

$$\overline{X^n} = E\left[X^n\right] = \sum_{x=-\infty}^{\infty} x^n \cdot f_X(x) = \sum_{x=-\infty}^{\infty} x^n \cdot \Pr(X = x)$$

Therefore, the mean or average is sometimes called the first moment.

Expected Mean Squared Value or Second Moment

The mean square value or second moment is

$$\overline{X^2} = E\left[X^2\right] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) \cdot dx$$

$$\overline{X^2} = E\left[X^2\right] = \sum_{x=-\infty}^{\infty} x^2 \cdot \Pr(X = x)$$

The second moment is related to the average "energy" or "power" in a signal, where the energy and power are defined as

$$E_x = \lim_{T \to \infty} \int_{-T/2}^{T/2} |x(t)|^2 \cdot dt \Rightarrow \int_{-\infty}^{\infty} |x(t)|^2 \cdot dt$$

$$P_x = \lim_{T \to \infty} \frac{1}{T} \cdot \int_{-T/2}^{T/2} |x(t)|^2 \cdot dt$$

Central Moments and Variance

The central moments are the moments of the difference between a random variable and its mean.

$$\overline{\left(X-\overline{X}\right)^n} = E\left[\left(X-\overline{X}\right)^n\right] = \int_{-\infty}^{\infty} \left(x-\overline{X}\right)^n \cdot f_X(x)\cdot dx$$

Notice that the first central moment is ***0*** …

$$\overline{\left(X-\overline{X}\right)^1} = E\left[\left(X-\overline{X}\right)\right] = \int_{-\infty}^{\infty} \left(x-\overline{X}\right)\cdot f_X(x)\cdot dx$$

$$\overline{\left(X-\overline{X}\right)^1} = \int_{-\infty}^{\infty} x\cdot f_X(x)\cdot dx - \overline{X}\cdot \int_{-\infty}^{\infty} f_X(x)\cdot dx$$

$$\overline{\left(X-\overline{X}\right)^1} = \overline{X} - \overline{X}\cdot 1 = 0$$

The second central moment is referred to as the **variance** of the random variable …

$$\sigma^2 = \overline{\left(X-\overline{X}\right)^2} = E\left[\left(X-\overline{X}\right)^2\right] = \int_{-\infty}^{\infty} \left(x-\overline{X}\right)^2 \cdot f_X(x)\cdot dx$$

The square root of the variance is defined as the standard deviation, σ

$$\sigma = \sqrt{\overline{\left(X-\overline{X}\right)^2}}$$

Note that the variance may also be computed as:

$$\sigma^2 = E\left[\left(X-\overline{X}\right)^2\right] = E\left[\left(X-\overline{X}\right)\cdot\left(X-\overline{X}\right)\right]$$

$$\sigma^2 = E\left[X^2 - 2\cdot X\cdot \overline{X} + \overline{X}^2\right]$$

$$\sigma^2 = E\left[X^2\right] - 2\cdot\overline{X}\cdot E[X] + \overline{X}^2$$

$$\sigma^2 = E\left[X^2\right] - 2\cdot\overline{X}\cdot\overline{X} + \overline{X}^2$$

$$\sigma^2 = E\left[X^2\right] - \overline{X}^2 = E\left[X^2\right] - E[X]^2$$

$$\sigma^2 = \overline{X^2} - \overline{X}^2$$

<u>*The variance is equal to the 2nd moment minus the square of the first moment..*</u>

Another estimate of future outcomes, is the value that <u>*minimizes the mean squared error*</u>.

$$\min E\left[(error)^2\right] = \min E\left[\left(X-\hat{x}\right)^2\right]$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Power and Energy Considerations

DC Power/Energy is related to the mean square $E[X]^2 = \mu^2 = \overline{X}^2$

AC Power/Energy is related to the variance $E\left[\left(X - \overline{X}\right)^2\right] = \sigma^2$

Notice that the 2nd moment has both AC and DC power terms ….

$$\overline{X^2} = \sigma^2 + \overline{X}^2 = \sigma^2 + \mu^2$$

$$E\left[X^2\right] = E\left[\left(X - \overline{X}\right)^2\right] + E[X]^2$$

Means and Variances of Defined density functions.

Table 4.3-1 Means, Variances and Mean-Square values for Common Continuous RVs

Family	pdf $f(x)$	Mean $\mu = E[X]$	Variance σ^2	Mean square $E[X^2]$
Uniform	$U(a,b)$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	$\frac{1}{3}(b^2+ab+a^2)$
Exponential	$\frac{1}{\mu}e^{-x/\mu}u(x)$	μ	μ^2	$2\mu^2$
Gaussian	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$\mu^2+\sigma^2$
Laplacian	$\frac{1}{\sqrt{2}\sigma}e^{-\frac{\sqrt{2}}{\sigma}\lvert x\rvert}$	0	σ^2	σ^2
Rayleigh	$\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}u(x)$	$\sqrt{\frac{\pi}{2}}\sigma$	$\left(2-\frac{\pi}{2}\right)\sigma^2$	$2\sigma^2$

Table 4.3-2 Means, Variances, and Mean-Square Values for Common Discrete RVs

Family	PMF $P(k)$	Mean $\mu = E[K]$	Variance σ^2	Mean square $E[K^2]$
Bernoulli	$P_B(k) = \begin{cases} 1, & p \\ 0, & q \end{cases} \; q \triangleq 1-p$	p	pq	p
Binomial	$b(k;n,p) = \binom{n}{k}p^k q^{n-k}$	np	npq	$(np)^2 + npq$
Geometric†	$\frac{1}{1+\mu}\left(\frac{\mu}{1+\mu}\right)^k u(k)$	μ	$\mu+\mu^2$	$\mu+2\mu^2$
Poisson	$\frac{\alpha^k}{k!}e^{-\alpha}u(k)$	α	α	$\alpha^2+\alpha$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Linearity of the expected values – a linear operator

The linearity allowing "linear operate" characteristics can be interpreted as allowing the expected value of a function to be the expected value of the internal elements of a function.

For example

$$E\left[\sum_{i-1}^{N} g_i(X)\right] = \int_{-\infty}^{\infty}\left[\sum_{i-1}^{N} g_i(X)\right]\cdot f_X(x)\cdot dx$$

But this can be expanded or rewritten as

$$E\left[\sum_{i-1}^{N} g_i(X)\right] = \int_{-\infty}^{\infty}\left[\sum_{i-1}^{N} g_i(X)\cdot f_X(x)\right]\cdot dx$$

The order of integration and summation can also be changed

$$E\left[\sum_{i-1}^{N} g_i(X)\right] = \sum_{i-1}^{N}\int_{-\infty}^{\infty}\left[g_i(X)\cdot f_X(x)\right]\cdot dx$$

So that we have

$$E\left[\sum_{i-1}^{N} g_i(X)\right] = \sum_{i-1}^{N} E\left[g_i(X)\right]$$

As another example

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[X+Y]\cdot f_{X,Y}(x,y)\cdot dx\cdot dy$$

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} X\cdot f_{X,Y}(x,y)\cdot dx\cdot dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Y\cdot f_{X,Y}(x,y)\cdot dx\cdot dy$$

$$E[X+Y] = \int_{-\infty}^{\infty} X\cdot f_X(x)\cdot dx + \int_{-\infty}^{\infty} Y\cdot f_Y(y)\cdot dy$$

$$E[X+Y] = E[X] + E[Y]$$

Examples

First Moment of exponential

for $f_X(x) = \lambda \cdot e^{-\lambda \cdot x} \cdot u(x)$

$$\overline{X} = E[X] = \int_0^\infty x \cdot \lambda \cdot e^{-\lambda \cdot x} \cdot dx$$

Integral Table Formula $\int x \cdot e^{a \cdot x} = \frac{e^{a \cdot x}}{a^2} \cdot (a \cdot x - 1)$

$$\overline{X} = E[X] = \lambda \cdot \frac{e^{-\lambda \cdot x}}{\lambda^2} \cdot (-\lambda \cdot x - 1)\bigg|_0^\infty$$

$$\overline{X} = E[X] = \left[\lambda \cdot \frac{e^{-\lambda \cdot \infty}}{\lambda^2} \cdot (-\lambda \cdot \infty - 1)\right] - \left[\lambda \cdot \frac{e^{-\lambda \cdot 0}}{\lambda^2} \cdot (-\lambda \cdot 0 - 1)\right]$$

$$\overline{X} = E[X] = 0 - \left[\lambda \cdot \frac{1}{\lambda^2} \cdot (-1)\right] = \frac{1}{\lambda}$$

Second Moment of exponential

for $f_X(x) = \lambda \cdot e^{-\lambda \cdot x} \cdot u(x)$

$$\overline{X^2} = E[X^2] = \int_0^\infty x^2 \cdot \lambda \cdot e^{-\lambda \cdot x} \cdot dx$$

Integral Table $\int x^2 \cdot e^{a \cdot x} = \frac{x^2 \cdot e^{a \cdot x}}{a} - \frac{2}{a} \cdot \int x \cdot e^{a \cdot x} = \frac{x^2 \cdot e^{a \cdot x}}{a} - \frac{2 \cdot x \cdot e^{a \cdot x}}{a^2} + \frac{2 \cdot e^{a \cdot x}}{a^3}$

$$\overline{X^2} = E[X^2] = \lambda \cdot e^{-\lambda \cdot x} \cdot \left(-\frac{x^2}{\lambda} - \frac{2 \cdot x}{\lambda^2} - \frac{2}{\lambda^3}\right)\bigg|_0^\infty$$

$$\overline{X^2} = E[X^2] = \left[\lambda \cdot e^{-\lambda \cdot \infty} \cdot \left(-\frac{\infty^2}{\lambda} - \frac{2 \cdot \infty}{\lambda^2} - \frac{2}{\lambda^3}\right)\right] - \left[\lambda \cdot e^{-\lambda \cdot 0} \cdot \left(-\frac{0^2}{\lambda} - \frac{2 \cdot 0}{\lambda^2} - \frac{2}{\lambda^3}\right)\right]$$

$$\overline{X^2} = E[X^2] = [0] - \left[-\frac{\lambda \cdot 2}{\lambda^3}\right] = \frac{2}{\lambda^2}$$

Example 4.1-7 Variance of Gaussian

$$E\left[(X-\mu)^2\right]=\int_{-\infty}^{\infty}(X-\mu)^2\cdot f_X(x)\cdot dx$$

$$f_X(x)=\frac{1}{\sqrt{2\pi\cdot\sigma^2}}\cdot\exp\left(\frac{-(x-\mu)^2}{2\cdot\sigma^2}\right),\ for\ -\infty<x<\infty$$

$$E\left[(X-\mu)^2\right]=\int_{-\infty}^{\infty}(X-\mu)^2\cdot\frac{1}{\sqrt{2\pi\cdot\sigma^2}}\cdot\exp\left(\frac{-(x-\mu)^2}{2\cdot\sigma^2}\right)\cdot dx$$

$$E\left[(X-\mu)^2\right]=\int_{-\infty}^{\infty}(X-\mu)^2\cdot\frac{1}{\sqrt{2\pi\cdot\sigma^2}}\cdot\exp\left(\frac{-(x-\mu)^2}{2\cdot\sigma^2}\right)\cdot dx$$

Letting $z=\frac{x-\mu}{\sigma}$ with $dz=\frac{dx}{\sigma}$

$$E\left[(X-\mu)^2\right]=\int_{-\infty}^{\infty}z^2\cdot\sigma^2\cdot\frac{1}{\sqrt{2\pi\cdot\sigma^2}}\cdot\exp\left(\frac{-z^2}{2}\right)\cdot\sigma\cdot dz$$

$$E\left[(X-\mu)^2\right]=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\int_{-\infty}^{\infty}z^2\cdot\exp\left(\frac{-z^2}{2}\right)\cdot dz$$

Integrating by parts: $\int v\cdot du=v\cdot u-\int u\cdot dv$

z	$z\cdot\exp\left(\frac{-z^2}{2}\right)$
1	$-\exp\left(\frac{-z^2}{2}\right)$

$$E\left[(X-\mu)^2\right]=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\left\{-z\cdot\exp\left(\frac{-z^2}{2}\right)\Bigg|_{-\infty}^{\infty}+\sqrt{2\pi}\cdot\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\cdot\exp\left(\frac{-z^2}{2}\right)\cdot dz\right\}$$

$$E\left[(X-\mu)^2\right]=\frac{\sigma^2}{\sqrt{2\pi}}\cdot\left\{-0+0+\sqrt{2\pi}\cdot 1\right\}=\sigma^2$$

Example 4.1-15 Geometric Distribution

$$P_X(x) = pmf_X(x) = (1-a)\cdot a^x, \qquad 0 \le x$$

Determine the expected value

$$E[X] = \mu = \sum_{x=0}^{\infty} x \cdot P_X(n)$$

$$E[X] = \sum_{x=0}^{\infty} x \cdot (1-a)\cdot a^x = (1-a)\cdot \sum_{x=0}^{\infty} x \cdot a^x$$

Note: $$\sum_{x=0}^{\infty} a^x = \frac{1}{1-a}, \quad for\ |a| < 1$$

and $$\frac{d}{da}\sum_{x=0}^{\infty} a^x = \sum_{x=0}^{\infty} x \cdot a^{x-1} = \frac{d}{da}\left(\frac{1}{1-a}\right) = \frac{1}{(1-a)^2}$$

$$E[X] = (1-a)\cdot a \cdot \sum_{x=0}^{\infty} x \cdot a^{x-1} = (1-a)\cdot a \cdot \frac{1}{(1-a)^2}$$

$$E[X] = \mu = \frac{a}{1-a}$$

This allows the mean value to be quickly found once “a” is known.

Determine the 2nd moment

$$E[X^2] = \sum_{x=0}^{\infty} x^2 \cdot P_X(n)$$

$$E[X^2] = \sum_{x=0}^{\infty} x^2 \cdot (1-a)\cdot a^x$$

Note: $$\frac{d^2}{da^2}\sum_{x=0}^{\infty} a^x = \sum_{x=0}^{\infty} x\cdot(x-1)\cdot a^{x-2} = \frac{d^2}{da^2}\left(\frac{1}{1-a}\right) = \frac{2}{(1-a)^3}$$

$$E[X^2] = (1-a)\cdot a^2 \cdot \sum_{x=0}^{\infty}\left[x\cdot(x-1)\cdot a^{x-2} + x \cdot a^{x-2}\right]$$

$$E[X^2] = (1-a)\cdot a^2 \cdot \sum_{x=0}^{\infty}\left[x\cdot(x-1)\cdot a^{x-2}\right] + (1-a)\cdot a \cdot \sum_{x=0}^{\infty} x \cdot a^{x-1}$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

$$E\left[X^2\right]=(1-a)\cdot a^2\cdot\frac{2}{(1-a)^3}+(1-a)\cdot a\cdot\frac{1}{(1-a)^2}$$

$$E\left[X^2\right]=\frac{2\cdot a^2}{(1-a)^2}+\frac{a}{1-a}=2\cdot\mu^2+\mu$$

Determine the variance

$$E\left[(X-\mu)^2\right]=E\left[X^2\right]-E[X]^2$$

$$E\left[(X-\mu)^2\right]=\left[\frac{2\cdot a^2}{(1-a)^2}+\frac{a}{1-a}\right]-\left[\frac{a}{1-a}\right]^2$$

$$E\left[(X-\mu)^2\right]=\frac{a^2}{(1-a)^2}+\frac{a}{1-a}=\mu^2+\mu$$

4.2 Conditional Expectations 232

Expected Values for Conditional Probability

The conditional expectation is defined as

$$E[X \mid B] \equiv \int_{-\infty}^{\infty} x \cdot f_{X|B}(x \mid B) \cdot dx$$

For discrete R.V.

$$E[X \mid B] \equiv \sum_i x_i \cdot \Pr_{X|B}(x_i \mid B) \equiv \sum_i x_i \cdot pmf_{X|B}(x_i \mid B)$$

As an "operator" the definition should be expected.

Example 4.2-1" Conditional expectation of a uniform R.V

$$f_X(x) = \frac{1}{x_1 - x_0}, \qquad x_0 < x \le x_1$$

$$F_X(x) = \frac{x - x_0}{x_1 - x_0}, \qquad x_0 < x \le x_1$$

The condition $B = \{X \ge a\}$

Therefore:

$$f_X(x \mid B) = \frac{f_X(x)}{1 - F_X(a)} = \frac{1}{x_1 - a}, \qquad a < x \le x_1$$

$$F_X(x \mid B) = \frac{F_X(x) - F_X(a)}{1 - F_X(a)} = \frac{x - a}{x_1 - a}, \quad a < x \le x_1$$

Using some numbers and performing expectations.

Assume that the RV was uniform from 0 to 100. What is the new expected value of X.

$$f_X(x) = \frac{1}{100 - 0} = \frac{1}{100}, \qquad 0 < x \le 100$$

$$F_X(x) = \frac{x - 0}{100 - 0} = \frac{x}{100}, \qquad 0 < x \le 100$$

$$E[X] = \int x \cdot f_X(x) \cdot dx = \int_0^{100} x \cdot \frac{1}{100} \cdot dx$$

$$E[X] = \frac{1}{100} \cdot \left. \frac{x^2}{2} \right|_0^{100} = \frac{1}{100} \cdot \left(\frac{100^2}{2} - \frac{0^2}{2} \right) = \frac{1}{100} \cdot \frac{100^2}{2} = \frac{100}{2} = 50$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

The conditional was above 65, what is the new expected value of X.

The condition $$B = \{X \geq 65\}$$

$$f_X(x \mid B) = \frac{f_X(x)}{1 - F_X(a)} = \frac{1}{100 - 65} = \frac{1}{35}, \qquad 65 < x \leq 100$$

$$E[X \mid B] = \int x \cdot f_{X|B}(x \mid B) \cdot dx = \int_{65}^{100} x \cdot \frac{1}{35} \cdot dx$$

$$E[X \mid B] = \frac{1}{35} \cdot \frac{x^2}{2}\bigg|_{65}^{100} = \frac{1}{35} \cdot \left(\frac{100^2}{2} - \frac{65^2}{2}\right) = \frac{1}{35} \cdot \frac{10000 - 4225}{2} = \frac{5775}{70} = 82.5$$

Expected Values for Joint Density functions and Conditional Probability

The expected values of joint density functions where multiple random variables are involved and may or may not have conditions occurs often.

- Public Health Considerations
- Effects of one variable on another for a statistical or probabilistic experiment.

Definition 4.2-3 The expected value of a conditional probability.

For the joint density function given as: $f_{X,Y}(x, y)$

We want to now the expected value

$$E[Y \mid X = x]$$

We know from before

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

$$E[Y \mid X = x] = \int y \cdot f_{X,Y}(y \mid X = x) \cdot dy$$

Usefulness or application is in the expected value of Y

$$E[Y] = \int \int y \cdot f_{X,Y}(x, y) \cdot dy \cdot dx$$

Application

$$E[Y]=\int\int y\cdot f_{X,Y}(x,y)\cdot dy\cdot dx$$

But $$f(x,y)=f(x\mid Y=y)\cdot f_Y(y)=f(y\mid X=x)\cdot f_X(x)$$

$$E[Y]=\int\int y\cdot f(y\mid X=x)\cdot f_X(x)\cdot dy\cdot dx$$

$$E[Y]=\int\left[\int y\cdot f(y\mid X=x)\cdot dy\right]\cdot f_X(x)\cdot dx$$

$$E[Y]=\int E[Y\mid X=x]\cdot f_X(x)\cdot dx$$

This may or may not seem like a logical result ….

Properties of Conditional Expectations:

Property I: Expected values of a conditional expected values:

$$E[Y] = E[E[Y \mid X]]$$

Derivation

$$E[Y] = \int y \cdot f_Y(y) \cdot dy$$

$$f_Y(y) = \int f_{X,Y}(x,y) \cdot dx$$

$$f(x,y) = f(x \mid Y = y) \cdot f_Y(y) = f(y \mid X = x) \cdot f_X(x)$$

$$f_Y(y) = \int f(y \mid X = x) \cdot f_X(x) \cdot dx$$

$$E[Y] = \int y \cdot \left[\int f(y \mid X = x) \cdot f_X(x) \cdot dx \right] \cdot dy$$

$$E[Y] = \int \left[\int y \cdot f(y \mid X = x) \cdot dy \right] \cdot f_X(x) \cdot dx$$

$$E[Y] = \int E[Y \mid X] \cdot f_X(x) \cdot dx = E[E[Y \mid X]]$$

Property II: If X and Y are independent (X should not matters)

$$E[Y] = E[Y \mid X]$$

Derivation

$$f(x,y) = f(x \mid Y = y) \cdot f_Y(y) = f(y \mid X = x) \cdot f_X(x)$$

$$f(x,y) = f_Y(y) \cdot f_X(x) = f(x \mid Y = y) \cdot f_Y(y) = f(y \mid X = x) \cdot f_X(x)$$

$$f_Y(y) \cdot f_X(x) = f(y \mid X = x) \cdot f_X(x)$$

$$f_Y(y) = f(y \mid X = x)$$

$$E[Y] = \int y \cdot f_Y(y) \cdot dy = \int y \cdot f(y \mid X = x) \cdot dy$$

$$E[Y] = E[Y \mid X]$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Property III: A conditional chain rule …

$$E[Z \mid X] = E[E[Z \mid X, Y] \mid X]$$

I hate the textbook proof !

$$E[Z \mid X] = \int z \cdot f_{Z|X}(z \mid X = x) \cdot dz$$

$$E[Z \mid X, Y] = \int z \cdot f_{Z|X,Y}(z \mid x, y) \cdot dz$$

Taking the expected value, only X,Y remain

$$E[E[Z \mid X, Y]] = \iint \left[\int z \cdot f_{Z|X,Y}(z \mid x, y) \cdot dz \right] \cdot f_{X,Y}(x, y) \cdot dx \cdot dy$$

But

$$E[E[Z \mid X, Y] \mid X] = \int \left[\int z \cdot f_{Z|X,Y}(z \mid x, y) \cdot dz \right] \cdot f_{Y|X}(y \mid x) \cdot dy$$

$$E[E[Z \mid X, Y] \mid X] = \int z \cdot \left[\int f_{Y|X}(y \mid x) \cdot dy \right] \cdot f_{Z|X,Y}(z \mid x, y) \cdot dz$$

Based on this equation, all values of y have been considered. Therefore,

$$E[E[Z \mid X, Y] \mid X] = \int z \cdot f_{Z|X}(z \mid X) \cdot dz = E[Z \mid X]$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

4.3 Moments of Random Variables 242

Example 4.3-1 Binomial R.V.

$$P_X(x) = pmf_X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Determine the expected value

$$E[X] = \mu = \sum_{x=0}^{n} x \cdot P_X(n)$$

$$E[X] = \sum_{x=0}^{n} x \cdot \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Proof based on Wikipedia https://en.wikipedia.org/wiki/Binomial_distribution

$$E[X] = \sum_{x=0}^{n} x \cdot \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot (1-p)^{n-x} = \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)! \cdot x!} \cdot p^x \cdot (1-p)^{n-x}$$

$$E[X] = \sum_{x=1}^{n} \frac{n!}{(n-x)! \cdot (x-1)!} \cdot p^x \cdot (1-p)^{n-x}$$

$$E[X] = \sum_{x=1}^{n} \frac{n \cdot (n-1)!}{((n-1)-(x-1))! \cdot (x-1)!} \cdot p \cdot p^{x-1} \cdot (1-p)^{(n-1)-(x-1)}$$

$$E[X] = n \cdot p \cdot \sum_{x=1}^{n} \frac{(n-1)!}{((n-1)-(x-1))! \cdot (x-1)!} \cdot p^{x-1} \cdot (1-p)^{(n-1)-(x-1)}$$

$$y = x-1$$

$$E[X] = n \cdot p \cdot \sum_{y=0}^{n-1} \frac{(n-1)!}{((n-1)-y)! \cdot y!} \cdot p^y \cdot (1-p)^{(n-1)-y}$$

But
$$\sum_{y=0}^{m} \frac{m!}{(m-y)! \cdot y!} \cdot p^y \cdot (1-p)^{m-y} = 1$$

$$E[X] = n \cdot p$$

Determine the 2nd moment

$$E[X^2] = \sum_{x=0}^{\infty} x^2 \cdot P_X(n)$$

$$E[X^2] = \sum_{x=0}^{n} [x-1+1] \cdot \frac{n!}{(n-x)! \cdot (x-1)!} \cdot p^x \cdot (1-p)^{n-x}$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

$$E\left[X^2\right] = \sum_{x=1}^{n} (x-1) \cdot \frac{n!}{(n-x)! \cdot (x-1)!} \cdot p^x \cdot (1-p)^{n-x} + \sum_{x=1}^{n} \frac{n!}{(n-x)! \cdot (x-1)!} \cdot p^x \cdot (1-p)^{n-x}$$

The second term was previously computed

$$E\left[X^2\right] = \sum_{x=2}^{n} \frac{n!}{(n-x)! \cdot (x-2)!} \cdot p^x \cdot (1-p)^{n-x} + n \cdot p$$

$$E\left[X^2\right] = n \cdot p + \sum_{x=2}^{n} \frac{(n-2)! \cdot n \cdot (n-1)}{((n-2)-(x-2))! \cdot (x-2)!} \cdot p^2 \cdot p^{x-2} \cdot (1-p)^{(n-2)-(x-2)} +$$

$$E\left[X^2\right] = n \cdot p + n \cdot (n-1) \cdot p^2 \cdot \sum_{x=2}^{n} \frac{(n-2)!}{((n-2)-(x-2))! \cdot (x-2)!} \cdot p^{x-2} \cdot (1-p)^{(n-2)-(x-2)} +$$

$$E\left[X^2\right] = n \cdot p + n \cdot (n-1) \cdot p^2 \cdot 1$$

$$E\left[X^2\right] = n \cdot (n-1) \cdot p^2 + n \cdot p$$

Determine the variance

$$E\left[(X-\mu)^2\right] = E\left[X^2\right] - E[X]^2$$

$$E\left[(X-\mu)^2\right] = n \cdot (n-1) \cdot p^2 + n \cdot p - [n \cdot p]^2$$

$$E\left[(X-\mu)^2\right] = n^2 p^2 - n \cdot p^2 + n \cdot p - n^2 p^2 = -n \cdot p^2 + n \cdot p$$

$$E\left[(X-\mu)^2\right] = n \cdot p \cdot (1-p) = n \cdot p \cdot q$$

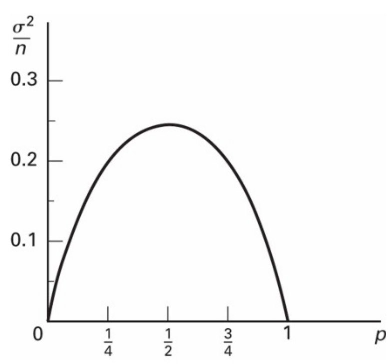


Figure 4.3-1 Variance of a binomial RV versus p.

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Joint Moments 246

When we have multiple random variables, additional moments can be defined.

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

All expected values may be computed using the Joint pdf. There are some "new" relationships.

Correlation and Covariance between Random Variables

The definition of correlation was given as

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

But most of the time, we are not interested in products of mean values (observed when X and Y are independent) but what results when they are removed prior to the computation. Developing values where the random variable means have been extracted, is defined as computing the ***covariance***

$$COV[X,Y] = E[(X-E[X])\cdot(Y-E[Y])] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)\cdot(y-\mu_Y)\cdot f(x,y)\cdot dx\cdot dy$$

This gives rise to another factor, when the random variable means and variances are used to normalize the factors or correlation/covariance computation. For example, the following definition – ***correlation coefficient*** based on the normalized covariance

$$\rho = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\cdot\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x-\mu_X}{\sigma_X}\right)\cdot\left(\frac{y-\mu_Y}{\sigma_Y}\right)\cdot f(x,y)\cdot dx\cdot dy$$

$$\rho = \frac{COV(X,Y)}{\sigma_X\cdot\sigma_Y}$$

Also

$$\rho = \frac{E[X\cdot Y]-\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y}$$

$$E[X\cdot Y] = \rho\cdot\sigma_X\cdot\sigma_Y+\mu_X\cdot\mu_Y$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

The short derivation

$$\rho = E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\cdot\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right] = E\left[\frac{X\cdot Y-\mu_X\cdot Y-\mu_Y\cdot X+\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y}\right]$$

The expected value is a linear operator … constants remain constants and sums are sums …

$$\rho = \frac{E[X\cdot Y]-\mu_X\cdot E[Y]-\mu_Y\cdot E[X]+\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y}$$

$$\rho = \frac{E[X\cdot Y]-\mu_X\cdot\mu_Y-\mu_Y\cdot\mu_X+\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y}$$

$$\rho = \frac{E[X\cdot Y]-\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y}$$

Properties of Uncorrelated Random Variables 248

For "standardized" random variables, the correlation coefficient can be solved for as the correlation value.

$$\rho = \frac{E[x\cdot y]-\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y} = \frac{E[x\cdot y]-0\cdot 0}{1\cdot 1} = E[x\cdot y]$$

For either X or Y a zero mean variable,

$$\rho\frac{E[x\cdot y]-0}{\sigma_X\cdot\sigma_Y} = \frac{E[X\cdot Y]}{\sigma_X\cdot\sigma_Y}$$

For independent random variables …

$$E[X\cdot Y] = E[X]\cdot E[Y] = \mu_X\cdot\mu_Y$$

$$\rho = \frac{E[X\cdot Y]-\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y} = \frac{\mu_X\cdot\mu_Y-\mu_X\cdot\mu_Y}{\sigma_X\cdot\sigma_Y} = 0$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Mean and Variance of the Sum of two R.V.

For X and Y uncorrelated … letting Z=X+Y

$$VAR[Z] = E\left[(Z-\mu_Z)^2\right] = E\left[(X+Y-\mu_X-\mu_Y)^2\right]$$

Note, Linear Op: $E[Z] = \mu_Z = E[X+Y] = E[X] + E[Y] = \mu_X + \mu_Y$

$$VAR[Z] = E\left[(X-\mu_X+Y-\mu_Y)^2\right] = E\left[(X-\mu_X)^2 + 2\cdot(X-\mu_X)\cdot(Y-\mu_Y) + (Y-\mu_Y)^2\right]$$

$$VAR[Z] = E\left[(X-\mu_X)^2\right] + E\left[(Y-\mu_Y)^2\right] + 2\cdot E\left[(X-\mu_X)\cdot(Y-\mu_Y)\right]$$

$$VAR[Z] = VAR[X] + 2\cdot COV(X,Y) + VAR[Y]$$

$$VAR[Z] = {\sigma_Z}^2 = {\sigma_X}^2 + 2\cdot\rho\cdot\sigma_X\cdot\sigma_Y + {\sigma_Y}^2$$

For uncorrelated X and Y

$$VAR[Z] = {\sigma_Z}^2 = {\sigma_X}^2 + {\sigma_Y}^2$$

A special note

Independent random variables are uncorrelated.

$$E[X \cdot Y] = E[X] \cdot E[Y] = \mu_X \cdot \mu_Y$$

$$\rho = \frac{E[X \cdot Y] - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y} = \frac{\mu_X \cdot \mu_Y - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y} = 0$$

However, uncorrelated random variables are not necessarily independent.

Example 4.3-5: Given

$P_{X,Y}$	$x_1 = -1$	$x_2 = 0$	$x_3 = 1$
$y_1 = 0$	0	$1/3$	0
$y_2 = 1$	$1/3$	0	$1/3$

$$P_X(x_i) = \sum_{j=1}^{2} P_{X,Y}(x_i, y_j) \qquad\qquad P_X(x_i) = \sum_{j=1}^{2} P_{X,Y}(x_i, y_j) = \frac{1}{3}$$

$$P_X(x_1) = P_X(x_2) = P_X(x_3) = \frac{1}{3}$$

$$P_Y(y_i) = \sum_{j=1}^{3} P_{X,Y}(x_j, y_i) \rightarrow P_Y(y_0) = \frac{1}{3}, P_Y(y_1) = \frac{2}{3}$$

Note, not independent $\quad P_{X,Y}(x_j, y_i) \neq P_X(x_j) \cdot P_Y(y_i)$

$$E[X] = \sum_{i=1}^{3} x_i \cdot P_X(x_i) = -1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$

$$E[Y] = \sum_{i=1}^{2} y_i \cdot P_Y(y_i) = 0\frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

$$E[X \cdot Y] = \sum_{i=1}^{2} \sum_{j=1}^{3} x_j \cdot y_i \cdot P_{X,Y}(x_j, y_i)$$

$$= -1 \cdot 0 \cdot 0 + 0 \cdot 0 \cdot \frac{1}{3} + 1 \cdot 0 \cdot 0 + -1 \cdot 1 \cdot \frac{1}{3} + 0 \cdot 1 \cdot 0 + 1 \cdot 1 \cdot \frac{1}{3} = 0$$

Therefore $\quad COV(X, Y) = 0$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Example 4.3-4 Linear Prediction – mean square error

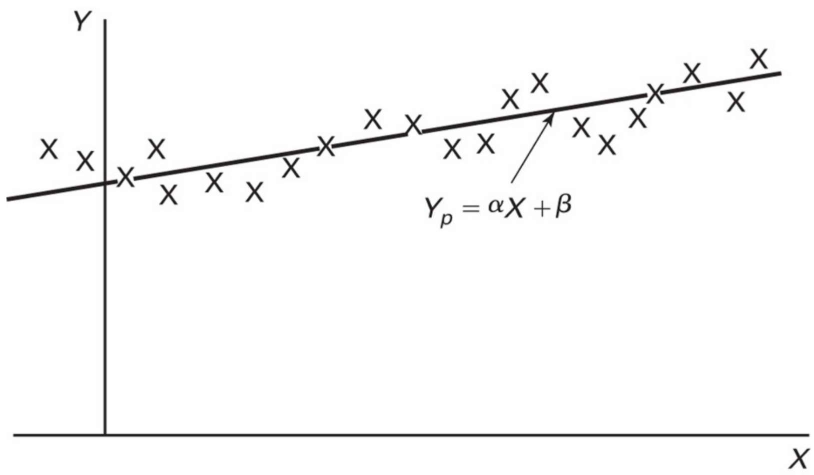


Figure 4.3-2 Pairwise observations on (*X, Y*) constitute a scatter diagram. The relationship between *X* and *Y* is approximated with a straight line.

Under a linear assumption of the relationship between X and Y

$$\hat{y} = a \cdot X + b$$

An error can be defined as

$$\varepsilon = Y - \hat{y} = Y - a \cdot X - b$$

We wish to form the variance of the error

$$E\left[\varepsilon^2\right] = E\left[(Y - \hat{y})^2\right] = E\left[(Y - a \cdot X - b)^2\right]$$

$$E\left[\varepsilon^2\right] = E\left[Y^2 - 2 \cdot Y \cdot (a \cdot X + b) + (a \cdot X + b)^2\right]$$

$$E\left[\varepsilon^2\right] = E\left[Y^2\right] - 2 \cdot E\left[Y \cdot (a \cdot X + b)\right] + E\left[(a \cdot X + b)^2\right]$$

$$E\left[\varepsilon^2\right] = E\left[Y^2\right] - 2 \cdot a \cdot E\left[Y \cdot X\right] - 2 \cdot b \cdot E[Y] + a \cdot^2 E\left[X^2\right] + 2 \cdot a \cdot b \cdot E[X] + b^2$$

$$E\left[\varepsilon^2\right] = E\left[Y^2\right] - 2 \cdot a \cdot E\left[Y \cdot X\right] - 2 \cdot b \cdot \mu_Y + a \cdot^2 E\left[X^2\right] + 2 \cdot a \cdot b \cdot \mu_X + b^2$$

Minimization says to take the derivative in a and set the derivative to zero and then derivative in b and set its derivative to zero.

Taking partial derivatives

$$\frac{\partial}{\partial a} E\left[\varepsilon^2\right] = -2 \cdot E[Y \cdot X] + 2 \cdot a \cdot E\left[X^2\right] + 2 \cdot b \cdot \mu_X = 0$$

$$2 \cdot a \cdot E\left[X^2\right] = 2 \cdot E[Y \cdot X] - 2 \cdot b \cdot \mu_X$$

Equation #1: $$2 \cdot a \cdot E\left[X^2\right] = E[Y \cdot X] - b \cdot \mu_X$$

$$\frac{\partial}{\partial b} E\left[\varepsilon^2\right] = -2 \cdot \mu_Y + 2 \cdot a \cdot \mu_X + 2 \cdot b = 0$$

$$2 \cdot b = 2 \cdot \mu_Y - 2 \cdot a \cdot \mu_X$$

Equation #2: $$b = \mu_Y - a \cdot \mu_X$$

Finding a (substitute b in #2 into #1)

$$a \cdot E\left[X^2\right] = E[Y \cdot X] - (\mu_Y - a \cdot \mu_X) \cdot \mu_X$$

$$a \cdot E\left[X^2\right] = E[Y \cdot X] - \mu_Y \cdot \mu_X + a \cdot {\mu_X}^2$$

$$a \cdot \left(E\left[X^2\right] - {\mu_X}^2\right) = E[Y \cdot X] - \mu_Y \cdot \mu_X$$

$$a = \frac{E[Y \cdot X] - \mu_Y \cdot \mu_X}{E\left[X^2\right] - {\mu_X}^2} = \frac{\rho \cdot \sigma_X \cdot \sigma_Y}{{\sigma_X}^2} = \frac{\rho \cdot \sigma_Y}{\sigma_X}$$

Finding b (using a and #2)

$$b = \mu_Y - a \cdot \mu_X = \mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X$$

The linear predictor becomes

$$\hat{y} = a \cdot X + b$$

$$\hat{y} = \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot X + \mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X = \mu_Y + \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (X - \mu_X)$$

Or

$$\hat{y} - \mu_Y = \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (X - \mu_X)$$

and the error is

$$\varepsilon = Y - \hat{y} = Y - \mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (X - \mu_X)$$

Finally, determining the "minimum" mean square error

$$E\left[\varepsilon^2\right] = E\left[Y^2\right] - 2 \cdot a \cdot E[Y \cdot X] - 2 \cdot b \cdot \mu_Y + a \cdot^2 E\left[X^2\right] + 2 \cdot a \cdot b \cdot \mu_X + b^2$$

$$E\left[\varepsilon^2\right] = E\left[Y^2\right] - 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot E[Y \cdot X] - 2 \cdot \left(\mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X\right) \cdot \mu_Y + \left(\frac{\rho \cdot \sigma_Y}{\sigma_X}\right)^2 E\left[X^2\right]$$

$$+ 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \left(\mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X\right) \cdot \mu_X + \left(\mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X\right)^2$$

$$E\left[\varepsilon^2\right] = {\sigma_Y}^2 + {\mu_Y}^2 - 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (\rho \cdot \sigma_X \cdot \sigma_Y + \mu_X \cdot \mu_Y) - 2 \cdot \left(\mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X\right) \cdot \mu_Y$$

$$+ \left(\frac{\rho \cdot \sigma_Y}{\sigma_X}\right)^2 \left({\sigma_X}^2 + {\mu_X}^2\right) + 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \left(\mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X\right) \cdot \mu_X + \left(\mu_Y - \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X\right)^2$$

$$E\left[\varepsilon^2\right] = {\sigma_Y}^2 + {\mu_Y}^2 - 2 \cdot \rho^2 \cdot {\sigma_Y}^2 - 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X \cdot \mu_Y - 2 \cdot {\mu_Y}^2 + 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_X \cdot \mu_Y$$

$$+ \rho^2 \cdot {\sigma_Y}^2 + \left(\frac{\rho^2 \cdot {\sigma_Y}^2}{{\sigma_X}^2}\right) \cdot {\mu_X}^2 + 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_Y \cdot \mu_X - 2 \cdot \frac{\rho^2 \cdot {\sigma_Y}^2}{{\sigma_X}^2} \cdot {\mu_X}^2$$

$$+ {\mu_Y}^2 - 2 \cdot \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot \mu_Y \cdot \mu_X + \left(\frac{\rho^2 \cdot {\sigma_Y}^2}{{\sigma_X}^2}\right) \cdot {\mu_X}^2$$

Deleting where possible leaves,

$$E\left[\varepsilon^2\right] = {\sigma_Y}^2 - \rho^2 \cdot {\sigma_Y}^2 = {\sigma_Y}^2 \cdot \left(1 - \rho^2\right)$$

Things to notice …. see the following page.

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Things to notice …. *the linear predictive fit is based on the probability values computed*.

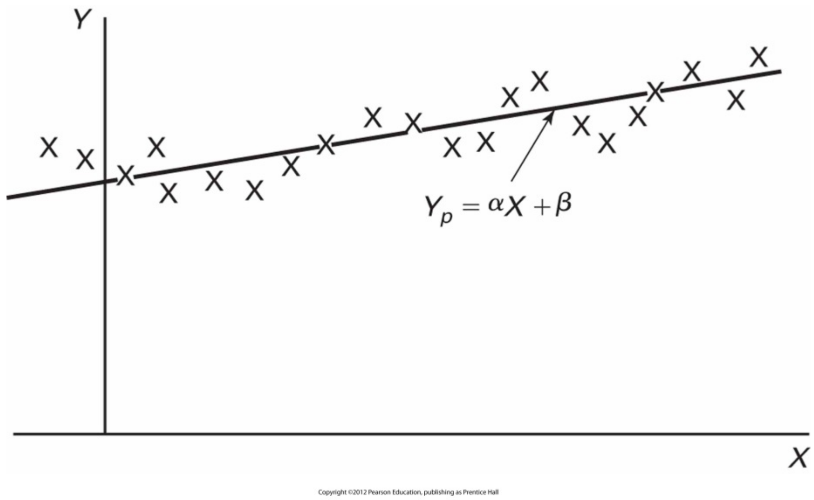


Figure 4.3-2 Pairwise observations on (*X, Y*) constitute a scatter diagram. The relationship between *X* and *Y* is approximated with a straight line.

$$\hat{y} - \mu_Y = \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (X - \mu_X)$$

$$E[\varepsilon^2] = {\sigma_Y}^2 - \rho^2 \cdot {\sigma_Y}^2 = {\sigma_Y}^2 \cdot (1 - \rho^2)$$

Meaning of the correlation coefficients … for a linear fit, correlation of some type is expected!

If $\rho = 0$:

$$\hat{y} = \mu_Y + \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (X - \mu_X) = \mu_Y$$
$$E[\varepsilon^2] = {\sigma_Y}^2 \cdot (1 - \rho^2) = {\sigma_Y}^2$$

If $\rho = \pm 1$:

$$\hat{y} = \mu_Y + \frac{\rho \cdot \sigma_Y}{\sigma_X} \cdot (X - \mu_X) = \mu_Y \pm \frac{\sigma_Y}{\sigma_X} \cdot (X - \mu_X)$$
$$\frac{\hat{y} - \mu_Y}{\sigma_Y} = \pm \frac{X - \mu_X}{\sigma_X}$$

$$E[\varepsilon^2] = {\sigma_Y}^2 \cdot (1 - \rho^2) = 0$$

Y is known one X is known, there for the estimation error must be zero!

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

If two R.V. are jointly Gaussian

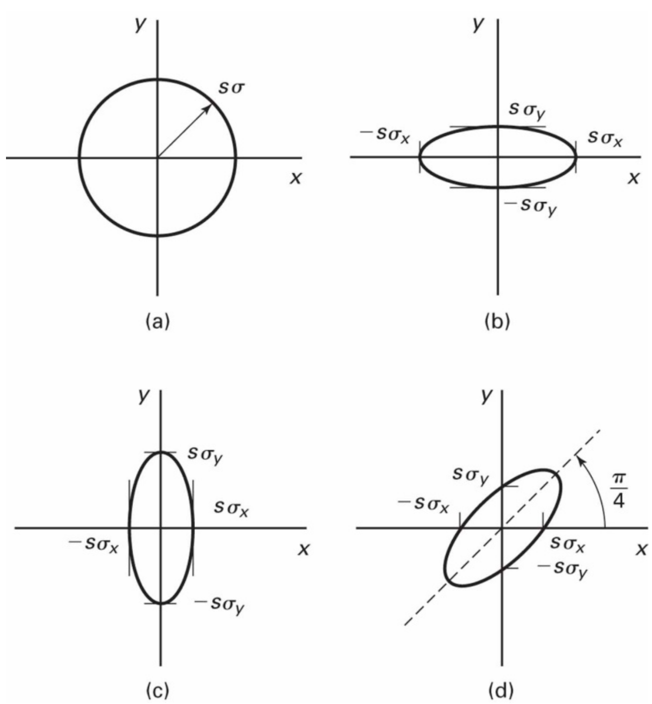
$$f_{X,Y}(x,y)=\frac{\exp\left[\frac{-1}{2\cdot\left(1-\rho^2\right)}\cdot\left\{\frac{(x-\mu_X)^2}{\sigma_X^2}-2\cdot\rho\cdot\frac{(x-\mu_X)\cdot(y-\mu_Y)}{\sigma_X\cdot\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right\}\right]}{2\pi\cdot\sigma_X\cdot\sigma_Y\cdot\sqrt{1-\rho^2}}$$

If $\rho=0$:

$$f_{X,Y}(x,y)=\frac{\exp\left[\frac{-(x-\mu_X)^2}{2\cdot\sigma_X^2}+\frac{-(y-\mu_Y)^2}{2\cdot\sigma_Y^2}\right]}{2\pi\cdot\sigma_X\cdot\sigma_Y}=\frac{\exp\left[\frac{-(x-\mu_X)^2}{2\cdot\sigma_X^2}\right]}{\sqrt{2\pi}\cdot\sigma_X}\cdot\frac{\exp\left[\frac{-(y-\mu_Y)^2}{2\cdot\sigma_Y^2}\right]}{\sqrt{2\pi}\cdot\sigma_Y}$$

Visualizing Joint Gaussians …

Figure 4.3-4 Contours of constant density for the joint normal (X = Y = 0): (a) $\sigma_X = \sigma_Y$, $\rho = 0$; (b) $\sigma_X > \sigma_Y$, $\rho=0$; (c) $\sigma_X < \sigma_Y$, $\rho=0$; (d) $\sigma_X = \sigma_Y$; $\rho>0$.



Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

There are a number of probability relationships that bound aspects of engineering problems. They are typically based on moments, particularly the mean and variance.
This is the first.

The Chebyshev inequality furnishes a bound on the probability of how much an R.V. can deviate from its mean value.

Chebyshev inequality Theorem 4.4-1

Let X be an arbitrary R.V. with known mean and variance. Then for any $\delta > 0$

$$P\left[\left|X-\mu_X\right| \ge \delta\right] \le \frac{{\sigma_X}^2}{\delta^2}$$

Derivation

$$\sigma^2 = \overline{\left(X-\overline{X}\right)^2} = E\left[\left(X-\overline{X}\right)^2\right] = \int_{-\infty}^{\infty} \left(x-\overline{X}\right)^2 \cdot f_X(x)\cdot dx$$

Then

$$\sigma^2 = \int_{-\infty}^{\infty} \left(x-\overline{X}\right)^2 \cdot f_X(x)\cdot dx \ge \int_{\left|x-\overline{X}\right|\ge\delta} \left(x-\overline{X}\right)^2 \cdot f_X(x)\cdot dx$$

and

$$\sigma^2 \ge \int_{\left|x-\overline{X}\right|\ge\delta} \left(x-\overline{X}\right)^2 \cdot f_X(x)\cdot dx \ge \int_{\left|x-\overline{X}\right|\ge\delta} (\delta)^2 \cdot f_X(x)\cdot dx = \delta^2 \cdot P\left[\left|x-\overline{X}\right| \ge \delta\right]$$

Results #1:

$$\frac{\sigma^2}{\delta^2} \ge P\left[\left|x-\overline{X}\right| \ge \delta\right]$$

If we also consider the complement of the probability described,

$$P\left[\left|x-\overline{X}\right| \ge \delta\right] + P\left[\left|x-\overline{X}\right| < \delta\right] = 1$$

and using the complement

$$\frac{\sigma^2}{\delta^2} \ge 1 - P\left[\left|x-\overline{X}\right| < \delta\right]$$

Therefore

Results #2:

$$P\left[\left|x-\overline{X}\right| < \delta\right] \ge 1 - \frac{\sigma^2}{\delta^2}$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

It may be convenient to define the delta function in terms of a multiple of the standard deviation.

$$\delta = k \cdot \sigma$$

Then the Chebyshev inequality becomes

$$P\left[\left|x-\overline{X}\right| \geq k \cdot \sigma\right] \leq \frac{1}{k^2}$$

$$P\left[\left|x-\overline{X}\right| < k \cdot \sigma\right] \geq 1-\frac{1}{k^2}$$

Example 4.4-1

Deviation from the mean for a Normal R.V.

The Gaussian Normal CDF is

$$\Phi_X\left(z=\frac{x-\overline{X}}{\sigma}\right)=\int_{v=-\infty}^{z}\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(\frac{-(v-\mu)^2}{2\cdot\sigma^2}\right)\cdot dv=\frac{1}{2}+erf(z)$$

Therefore

$$P\left[\left|\frac{x-\overline{X}}{\sigma}\right|<k\right]=\int_{v=-k}^{k}\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(\frac{-(v-\mu)^2}{2\cdot\sigma^2}\right)\cdot dv=2\cdot erf(k)$$

$$P\left[\left|x-\overline{X}\right| < k \cdot \sigma\right]=2\cdot erf(k)\geq 1-\frac{1}{k^2}$$

and

$$P\left[\left|x-\overline{X}\right| \geq k \cdot \sigma\right]=1-2\cdot erf(k)\leq \frac{1}{k^2}$$

The following table compares the Chebyshev inequality to the above function.

Table 4.4-1

k	$P[\lvert X-\overline{X}\rvert < k\sigma]$	CB	$P[\lvert X-\overline{X}\rvert > k\sigma]$	CB
0	0	0	1	1
0.5	0.383	0	0.617	1
1.0	0.683	0	0.317	1
1.5	0.866	0.556	0.134	0.444
2.0	0.955	0.750	0.045	0.250
2.5	0.988	0.840	0.012	0.160
3.0	0.997	0.889	0.003	0.111

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

The Markov inequality focuses on the mean values and states. For a non-negative R.V.

$$f_X(x) = 0, \qquad for x < 0$$

Then

$$P[X \geq \delta] \leq \frac{E[X]}{\delta}$$

Derivation

$$E[X] = \int_0^\infty x \cdot f_X(x) \cdot dx \geq \int_\delta^\infty x \cdot f_X(x) \cdot dx \geq \int_\delta^\infty \delta \cdot f_X(x) \cdot dx = \delta \cdot P[X \geq \delta]$$

Therefore

$$P[X \geq \delta] \leq \frac{E[X]}{\delta}$$

Example 4.4-2 Bad Resistors

Resistors have a mean value of 1000 ohms. If all resistors are to be measured and those above 1500 ohms, discarded how many might you estimate would be discarded?

$$P[X \geq 1500] \leq \frac{1000}{1500} = \frac{2}{3}$$

Note: this is a "bound" not the exact value, the exact value could be expected to be smaller than 67% as the inequality suggests. .

The Schwarz Inequality 258

Schwarz Inequality considers the magnitude of the covariance of two R.C.

$\rho = \dfrac{COV(X,Y)}{\sigma_X \cdot \sigma_Y}$ equality will hold if and only if Y is a linear function of X (a correlation coefficient of +/-1).

The inequality can be written as

$$Cov[X,Y]^2 \le \sigma_X{}^2 \cdot \sigma_Y{}^2$$

or

$$|Cov[X,Y]| \le \sqrt{\sigma_X{}^2 \cdot \sigma_Y{}^2}$$

Derivation: The covariance definition leads to a straight forward recognition of this inequality

$$\rho = \frac{COV(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Therefore

$$Cov[X,Y]^2 = \rho^2 \cdot \sigma_X{}^2 \cdot \sigma_Y{}^2$$

with $-1 \le \rho \le 1$

You have already experienced the Schwartz Inequality in other setting …

For the inner product

$$(h,g) = \int_{-\infty}^{\infty} h(x) \cdot g(x)^* \cdot dx$$

$$|(h,g)| \le \|h\| \cdot \|g\|$$

You may also see the convolution form as

$$\left\| \int_{-\infty}^{\infty} h(x) \cdot g(x)^* \cdot dx \right\| \le \sqrt{\left\| \int_{-\infty}^{\infty} h(x) \cdot h(x)^* \cdot dx \right\| \cdot \left\| \int_{-\infty}^{\infty} g(x) \cdot g(x)^* \cdot dx \right\|}$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Law of Large Numbers

Now that we have discussed the Cheybyshev inequality, we can provide a proof of the Law of Large Numbers. The discussions here provides the condition that the sample mean converges to the ensemble mean … that is the statistical mean equals the R.V. ensemble mean.

Example 4.4-3. The sample mean equals the expected value mean

For a large enough number of samples, we say that

$$\hat{\mu}_X = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

If we take the expected value

$$E[\hat{\mu}_X] = E\left[\frac{1}{n} \cdot \sum_{i=1}^{n} X_i\right]$$

$$E[\hat{\mu}_X] = \frac{1}{n} \cdot E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n} \cdot \sum_{i=1}^{n} E[X_i]$$

$$E[\hat{\mu}_X] = \frac{1}{n} \cdot \sum_{i=1}^{n} \mu_X = \frac{1}{n} \cdot n \cdot \mu_X = \mu_X$$

and

$$Var[\hat{\mu}_X] = \frac{1}{n^2} \cdot Var\left[\sum_{i=1}^{n} X_i\right]$$

$$Var[\hat{\mu}_X] = \frac{1}{n^2} \cdot n \cdot \sigma_X{}^2$$

$$Var[\hat{\mu}_X] = \frac{1}{n} \cdot \sigma_X{}^2$$

Therefore, using the Chebyshev inequality we state that

$$P[|\hat{\mu}_X - \mu_X| \geq \delta] \leq \frac{\sigma_X{}^2}{n \cdot \delta^2}$$

Then for any fixed value of delta

$$\lim_{n\to\infty} P[|\hat{\mu}_X - \mu_X| \geq \delta] \leq \lim_{n\to\infty} \frac{\sigma_X{}^2}{n \cdot \delta^2}$$

and

$$\lim_{n\to\infty} P[|\hat{\mu}_X - \mu_X| \geq \delta] \leq 0$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

4.5 Moment-Generating Functions 261

The text is now moving into some advanced concepts that support mathematical derivation of higher order moments.

I have been exposed to problems where the 4th moment of a R.V. required as part of a solution. If you really like and are comfortable with Laplace and Fourier Transforms these approach provide solutions faster and more easily than more brute force integral approaches.

The moment generation function (MGF) is the two sided Laplace transform of the probability density function (pdf). If the MGF exists, there is a forward and inverse relationship between the MGF and the pdf. The MGF is defined bases on the expected value as

$$M_X(t) = E[\exp(t \cdot X)]$$

Therefore

$$M_X(t) = \int_{-\infty}^{\infty} f_X(x) \cdot \exp(t \cdot x) \cdot dx$$

If you like s better than t in your Laplace transforms …

$$M_X(s) = \int_{-\infty}^{\infty} f_X(x) \cdot \exp(s \cdot x) \cdot dx$$

For discrete R.V. we perform a discrete Laplace transform

$$M_X(s) = \sum_{i=-\infty}^{\infty} pmf_X(x_i) \cdot \exp(s \cdot x_i) = \sum_{i=-\infty}^{\infty} P_X(x_i) \cdot \exp(s \cdot x_i)$$

Why do we do this?

1. It enables a convenient computation of the higher order moments
2. It can be used to estimate $f_x(x)$ from experimental measurements of the moments
3. It can be used to solve problems involving the computation of the sums of R.V.
4. It is an important analytical instrument that can be used to demonstrate results and establish additional bounds (the Chernoff Bound and the Central Limit Theorem).

It enables a convenient computation of the higher order moments

Based on the definition

$$M_X(t) = E[\exp(t \cdot X)]$$

Perform the Taylor series expansion of the exponential

$$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$$

$$M_X(t) = E[\exp(t \cdot X)] = E\left[1 + \frac{(t \cdot X)}{1!} + \frac{(t \cdot X)^2}{2!} + \cdots + \frac{(t \cdot X)^n}{n!} + \cdots\right]$$

or

$$M_X(t) = E[\exp(t \cdot X)] = 1 + \frac{(t \cdot m_1)}{1!} + \frac{(t \cdot m_2)^2}{2!} + \cdots + \frac{(t \cdot m_n)^n}{n!} + \cdots$$

The m_i are the i^{th} moments of the density function!

So how would we solve for the moments? By taking derivatives and setting t=0!

$$\frac{\partial^k}{\partial t^k} M_X(t) = m_k + \frac{(t \cdot m_{k+1})^1}{1!} + \cdots + \frac{(t \cdot m_n)^{n-k}}{(n-k)!} + \cdots$$

by setting t=0

$$\left.\frac{\partial^k}{\partial t^k} M_X(t)\right|_{t=0} = M_X^{(k)}(0) = m_k$$

Solution done by performing a 2-sided Laplace Transform and differentiation!

Example 4.5-1: The MGF of a Gaussian

$$f_X(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-(x-\mu)^2}{2 \cdot \sigma^2} \right), \textit{ for } -\infty < x < \infty$$

MGF:
$$M_X(t) = \int_{-\infty}^{\infty} f_X(x) \cdot \exp(t \cdot x) \cdot dx$$

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-(x-\mu)^2}{2 \cdot \sigma^2} \right) \cdot \exp(t \cdot x) \cdot dx$$

When integrating Gaussians … form an integral of a correctly formed Gaussian function and equate it to 1.0.

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-\left(x^2 - 2 \cdot \mu \cdot x - \mu^2\right) + 2 \cdot \sigma^2 \cdot t \cdot x}{2 \cdot \sigma^2} \right) \cdot dx$$

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{x^2 - 2 \cdot \left(\mu + \sigma^2 \cdot t\right) \cdot x + \mu^2}{2 \cdot \sigma^2} \right) \cdot dx$$

$$M_X(t) = \exp\left(\frac{\left(\mu + \sigma^2 \cdot t\right)^2 - \mu^2}{2 \cdot \sigma^2} \right) \cdot$$
$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{x^2 - 2 \cdot \left(\mu + \sigma^2 \cdot t\right) \cdot x + \left(\mu + \sigma^2 \cdot t\right)^2}{2 \cdot \sigma^2} \right) \cdot dx$$

$$M_X(t) = \exp\left(\frac{\left(\mu + \sigma^2 \cdot t\right)^2 - \mu^2}{2 \cdot \sigma^2} \right) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{\left(x - \left(\mu + \sigma^2 \cdot t\right)\right)^2}{2 \cdot \sigma^2} \right) \cdot dx$$

The integral is now equal to 1.0. And we have

$$M_X(t) = \exp\left(\frac{\left(\mu^2 + 2 \cdot \sigma^2 \cdot t \cdot \mu + \sigma^4 \cdot t^2\right) - \mu^2}{2 \cdot \sigma^2} \right) = \exp\left(\frac{2 \cdot \sigma^2 \cdot t \cdot \mu + \sigma^4 \cdot t^2}{2 \cdot \sigma^2} \right)$$

$$M_X(t) = \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2} \right)$$

Now we can generate the moments of a Gaussian function.

The 1st Moment

$$\left.\frac{\partial}{\partial t} M_X(t)\right|_{t=0} = \frac{\partial}{\partial t} \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right) = \left(\mu + \sigma^2 \cdot t\right) \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial}{\partial t} M_X(t)\right|_{t=0} = \left(\mu + \sigma^2 \cdot 0\right) \cdot \exp\left(\mu \cdot 0 + \frac{\sigma^2 \cdot 0^2}{2}\right) = \mu \cdot 1 = \mu$$

The 2nd Moment

$$\left.\frac{\partial^2}{\partial t^2} M_X(t)\right|_{t=0} = \frac{\partial}{\partial t} \left(\mu + \sigma^2 \cdot t\right) \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial^2}{\partial t^2} M_X(t)\right|_{t=0} = \left(\mu + \sigma^2 \cdot t\right)^2 \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right) + \left(\sigma^2\right) \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial^2}{\partial t^2} M_X(t)\right|_{t=0} = \left(\mu + \sigma^2 \cdot 0\right)^2 \cdot 1 + \left(\sigma^2\right) \cdot 1 = \mu^2 + \sigma^2$$

The 3rd Moment

$$\left.\frac{\partial^3}{\partial t^3} M_X(t)\right|_{t=0} = \frac{\partial}{\partial t} \left[\left(\mu + \cdot \sigma^2 \cdot t\right)^2 + \sigma^2\right] \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial^3}{\partial t^3} M_X(t)\right|_{t=0} = \left[\left(\mu + \cdot \sigma^2 \cdot t\right)^3 + \sigma^2 \cdot \left(\mu + \cdot \sigma^2 \cdot t\right)\right] \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$
$$+ \left[2 \cdot \left(\mu + \cdot \sigma^2 \cdot t\right) \cdot \sigma^2\right] \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial^3}{\partial t^3} M_X(t)\right|_{t=0} = \left[\left(\mu + \cdot \sigma^2 \cdot 0\right)^3 + \sigma^2 \cdot \left(\mu + \cdot \sigma^2 \cdot 0\right)\right] \cdot 1 + \left[2 \cdot \left(\mu + \cdot \sigma^2 \cdot 0\right) \cdot \sigma^2\right] \cdot 1$$

$$\left.\frac{\partial^3}{\partial t^3} M_X(t)\right|_{t=0} = \mu^3 + 3 \cdot \mu \cdot \sigma^2$$

The 4th Moment

$$\left.\frac{\partial^4}{\partial t^4} M_X(t)\right|_{t=0} = \frac{\partial}{\partial t}\left[\left(\mu + \cdot\sigma^2 \cdot t\right)^3 + 3 \cdot \sigma^2 \cdot \left(\mu + \cdot\sigma^2 \cdot t\right)\right] \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial^4}{\partial t^4} M_X(t)\right|_{t=0} = \left[\left(\mu + \cdot\sigma^2 \cdot t\right)^4 + 3 \cdot \sigma^2 \cdot \left(\mu + \cdot\sigma^2 \cdot t\right)^2\right] \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$
$$+ \left[3 \cdot \left(\mu + \cdot\sigma^2 \cdot t\right)^2 \cdot \sigma^2 + 3 \cdot \sigma^2 \cdot \left(0 + \cdot\sigma^2\right)\right] \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\left.\frac{\partial^4}{\partial t^4} M_X(t)\right|_{t=0} = \left[\left(\mu + \cdot\sigma^2 \cdot 0\right)^4 + 3 \cdot \sigma^2 \cdot \left(\mu + \cdot\sigma^2 \cdot 0\right)^2\right] \cdot 1$$
$$+ \left[3 \cdot \left(\mu + \cdot\sigma^2 \cdot 0\right)^2 \cdot \sigma^2 + 3 \cdot \sigma^2 \cdot \left(0 + \cdot\sigma^2\right)\right] \cdot 1$$

$$\left.\frac{\partial^4}{\partial t^4} M_X(t)\right|_{t=0} = \mu^4 + 6 \cdot \sigma^2 \cdot \mu^2 + 3 \cdot \sigma^4$$

See: https://en.wikipedia.org/wiki/Normal_distribution

Additional useful examples:

Example 4.5-2 MGF of Binomial

$$pmf_X(k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

MGF:
$$M_X(t) = \int_{-\infty}^{\infty} f_X(x) \cdot \exp(t \cdot x) \cdot dx$$

$$M_X(t) = \sum_{n=0}^{n} \binom{n}{k} \cdot p^k \cdot q^{n-k} \cdot \exp(t \cdot k)$$

$$M_X(t) = \sum_{n=0}^{n} \binom{n}{k} \cdot \left(p \cdot \exp(t)\right)^k \cdot q^{n-k}$$

Magical math step … not really, but I haven’t done the derivation myself …

$$M_X(t) = \left(p \cdot \exp(t) + q\right)^n$$

The 1st Moment

$$\left.\frac{\partial}{\partial t}M_X(t)\right|_{t=0}=\frac{\partial}{\partial t}\left((p\cdot\exp(t)+q)^n\right)=n\cdot(p\cdot\exp(t)+q)^{n-1}\cdot p\cdot\exp(t)\Big|_{t=0}$$

$$\left.\frac{\partial}{\partial t}M_X(t)\right|_{t=0}=n\cdot(p+q)^{n-1}\cdot p=n\cdot p$$

$$\left.\frac{\partial}{\partial t}M_X(t)\right|_{t=0}=\left.\frac{\alpha\cdot(1-\alpha)\cdot\exp(t)}{(1-\alpha\cdot\exp(t))^2}\right|_{t=0}=\frac{\alpha\cdot(1-\alpha)\cdot 1}{(1-\alpha\cdot 1)^2}=\frac{\alpha}{1-\alpha}$$

The 2nd Moment

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0}=\frac{\partial}{\partial t}\left(n\cdot p\cdot\exp(t)\cdot(p\cdot\exp(t)+q)^{n-1}\right)\Big|_{t=0}$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0}=n\cdot p\cdot\exp(t)\cdot(p\cdot\exp(t)+q)^{n-1}\Big|_{t=0}$$
$$+n\cdot p\cdot\exp(t)\cdot(n-1)\cdot(p\cdot\exp(t)+q)^{n-2}\cdot p\cdot\exp(t)\Big|_{t=0}$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0}=n\cdot p\cdot(p+q)^{n-1}+n\cdot p^2\cdot(n-1)\cdot(p+q)^{n-2}$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0}=n\cdot p+n\cdot p^2\cdot(n-1)=(n\cdot p)^2+n\cdot p\cdot(1-p)$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0}=(n\cdot p)^2+n\cdot p\cdot q$$

Example 4.5-3 MGF of Geometric Distribution

$$pmf_X(n) = (1-\alpha)\cdot\alpha^n,\ for\ 0 < \alpha < 1$$

MGF:
$$M_X(t) = \int_{-\infty}^{\infty} f_X(x)\cdot\exp(t\cdot x)\cdot dx$$

$$M_X(t) = \sum_{n=0}^{\infty}(1-\alpha)\cdot\alpha^n\cdot\exp(t\cdot n)$$

$$M_X(t) = \frac{1-\alpha}{1-\alpha\cdot\exp(t)}$$

The 1st Moment

$$\left.\frac{\partial}{\partial t}M_X(t)\right|_{t=0} = \frac{\partial}{\partial t}\left(\frac{1-\alpha}{1-\alpha\cdot\exp(t)}\right) = \left.\left(-\frac{1-\alpha}{(1-\alpha\cdot\exp(t))^2}\right)\cdot(-\alpha\cdot\exp(t))\right|_{t=0}$$

$$\left.\frac{\partial}{\partial t}M_X(t)\right|_{t=0} = \left.\frac{\alpha\cdot(1-\alpha)\cdot\exp(t)}{(1-\alpha\cdot\exp(t))^2}\right|_{t=0} = \frac{\alpha\cdot(1-\alpha)\cdot 1}{(1-\alpha\cdot 1)^2} = \frac{\alpha}{1-\alpha}$$

The 2nd Moment

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0} = \frac{\partial}{\partial t}\left(\frac{\alpha\cdot(1-\alpha)\cdot\exp(t)}{(1-\alpha\cdot\exp(t))^2}\right)$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0} = \frac{\alpha\cdot(1-\alpha)\cdot\exp(t)}{(1-\alpha\cdot\exp(t))^2} - 2\cdot\frac{\alpha\cdot(1-\alpha)\cdot\exp(t)}{(1-\alpha\cdot\exp(t))^3}\cdot(-\alpha\cdot\exp(t))$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0} = \left.\frac{\alpha\cdot(1-\alpha)\cdot\exp(t)}{(1-\alpha\cdot\exp(t))^2}\right|_{t=0} + 2\cdot\left.\frac{\alpha^2\cdot(1-\alpha)\cdot\exp(2\cdot t)}{(1-\alpha\cdot\exp(t))^3}\right|_{t=0}$$

$$\left.\frac{\partial^2}{\partial t^2}M_X(t)\right|_{t=0} = \frac{\alpha\cdot}{1-\alpha} + \frac{2\cdot\alpha^2}{(1-\alpha)^2}$$

Checking the results from Table 4.3-2 All these results match, as expected.

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

4.6 Chernoff Bound 264

The Chernoff bound is based on concepts related to the MGF.

Looking at "the tail probability" for $P[X \geq a]$ where 'a' is a defined constant.

As a tail, we expect the density function to have greatly decreased in this region … like an exponential or Gaussian.

Then, we can state

$$P[X \geq a] = \int_{a}^{\infty} f_X(x) \cdot dx$$

$$P[X \geq a] = \int_{-\infty}^{\infty} f_X(x) \cdot u(x-a) \cdot dx$$

If the tail is such that $u(x-a) \leq \exp(t \cdot (x-a))$

Then

$$P[X \geq a] = \int_{-\infty}^{\infty} f_X(x) \cdot u(x-a) \cdot dx \leq \int_{-\infty}^{\infty} f_X(x) \cdot \exp(t \cdot (x-a)) \cdot dx$$

But then

$$P[X \geq a] \leq \exp(-a \cdot t) \cdot \int_{-\infty}^{\infty} f_X(x) \cdot \exp(x \cdot t) \cdot dx = \exp(-a \cdot t) \cdot M_X(t)$$

To minimize the bound, the value of t that provides the smallest value should be used. Therefore, differentiate with respect to t, find the value of t for the minimum and use the derived value for the bounds.

Example 4.6-1 Chernoff Bound to Gaussian

Let X be Gaussian and consider the bounds where a> E[X]

For the Gaussian

$$f_X(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(\frac{-(x-\mu)^2}{2 \cdot \sigma^2} \right), \textit{for } -\infty < x < \infty$$

$$M_X(t) = \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2} \right)$$

The Bound becomes

$$P[X \geq a] \leq \exp(-a \cdot t) \cdot M_X(t) = \exp(-a \cdot t) \cdot \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$P[X \geq a] \leq \exp\left((\mu - a) \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

To find the minimum value of t

$$\frac{d}{dt}\exp\left((\mu - a) \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right) = 0$$

$$\exp\left((\mu - a) \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right) \cdot ((\mu - a) + \sigma^2 \cdot t) = 0$$

$$(\mu - a) + \sigma^2 \cdot t = 0$$

$$t = \frac{a - \mu}{\sigma^2}$$

Therefore

$$P[X \geq a] \leq \exp\left((\mu - a) \cdot \frac{a - \mu}{\sigma^2} + \frac{\sigma^2}{2} \cdot \left(\frac{a - \mu}{\sigma^2}\right)^2\right)$$

$$P[X \geq a] \leq \exp\left(-\frac{(a - \mu)^2}{\sigma^2} + \frac{1}{2} \cdot \frac{(a - \mu)^2}{\sigma^2}\right)$$

$$P[X \geq a] \leq \exp\left(-\frac{1}{2} \cdot \frac{(a - \mu)^2}{\sigma^2}\right), \qquad for\ a > \mu$$

4.7 Characteristic Functions 266

The characteristic function is to the MGF as the Fourier transform is to the Laplace transform. Instead of t being a complex variable (similar to s), we set t = jw.

Therefore, the characteristic function is

$$\Phi_X(w) = E[\exp(j \cdot w \cdot X)]$$

$$\Phi_X(w) = \int_{-\infty}^{\infty} f_X(x) \cdot \exp(j \cdot w \cdot X) \cdot dx$$

For discrete RV

$$\Phi_X(w) = \sum_{i=-\infty}^{\infty} pmf_X(x_i) \cdot \exp(j \cdot w \cdot x_i) = \sum_{i=-\infty}^{\infty} P_X(x_i) \cdot \exp(j \cdot w \cdot x_i)$$

In general, the CF has similar properties to the MGF. In addition to those previously mentioned, the Fourier transform is very useful when performing time domain convolution.

For sums of R.V., Z=X+Y, as previously shown the density function is the convolution of the two density function in X and Y.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) \cdot dx = \int_{-\infty}^{\infty} f_X(z-y) \cdot f_Y(y) \cdot dy$$

But this can also be performed in the “frequency domain” by multiplication

$$\Phi_Z(w) = \Phi_X(w) \cdot \Phi_Y(w)$$

As a result it makes it much easier to contemplate

$$Z = X_1 + X_2 + \cdots + X_N$$

Which would become

$$\Phi_Z(w) = \Phi_{X_1}(w) \cdot \Phi_{X_2}(w) \cdot \ldots \cdot \Phi_{X_N}(w)$$

Example 4.7-2

For $Z = X_1 + X_2 + \cdots + X_N$, where the X_i are independent and identically distributed (IID) solve for the new distribution when X_i is a Gaussian normal R.V. (0 mean, unit variance).

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-x^2}{2}\right), \; for \; -\infty < x < \infty$$

$$M_G(t) = \exp\left(\mu \cdot t + \frac{\sigma^2 \cdot t^2}{2}\right)$$

$$\Phi_G(w) = \exp\left(j \cdot w \cdot \mu + \frac{\sigma^2 \cdot (j \cdot w)^2}{2}\right) = \exp\left(j \cdot w \cdot \mu - \frac{\sigma^2 \cdot w^2}{2}\right)$$

Then for zero mean, unit variance

$$\Phi_X(w) = \exp\left(-\frac{w^2}{2}\right)$$

and for IID R.V.

$$\Phi_Z(w) = \Phi_{X_1}(w) \cdot \Phi_{X_2}(w) \cdot \ldots \cdot \Phi_{X_N}(w)$$

$$\Phi_Z(w) = \left[\exp\left(-\frac{w^2}{2}\right)\right]^N = \exp\left(-\frac{N \cdot w^2}{2}\right)$$

$$\Phi_Z(w) = \exp\left(-\frac{N \cdot w^2}{2}\right) = \exp\left(-\frac{\left(\sqrt{N}\right)^2 \cdot w^2}{2 \cdot}\right)$$

The inverse transform (based on the forward transform with a variance) results in

$$f_Z(z) = \frac{1}{\sqrt{2\pi \cdot N}} \cdot \exp\left(\frac{-z^2}{2 \cdot N}\right), \; for \; -\infty < z < \infty$$

Moment Generation with the Characteristic Function

AS with the MGF, the CF can generate moments by differentiation.

$$\left.\frac{\partial^k}{\partial t^k} M_X(t)\right|_{t=0} = M_X^{(k)}(0) = m_k$$

$$\left.\frac{\partial^k}{\partial w^k} \Phi_X(w)\right|_{w=0} = \Phi_X^{(k)}(0) = j^k \cdot m_k$$

or

$$m_k = \frac{1}{j^k} \cdot \Phi_X^{(k)}(0)$$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Joint MGF and CF

Joint MGF and CRF Function can and are defined. As may be expected, they can be used to compute "cross-products" of random variables just as they generated moments.

The Joint MGF is defined as

$$M_{X_1,X_2,\cdots,X_N}(t_1,t_2,\cdots,t_N) = E[\exp(t_1 \cdot X_1 + t_2 \cdot X_2 + \ldots + t_N \cdot X_N)]$$

The Joint CF is defined as

$$\Phi_{X_1,X_2,\cdots,X_N}(w_1,w_2,\cdots,w_N) = E[\exp(j \cdot w_1 \cdot X_1 + j \cdot w_2 \cdot X_2 + \ldots + j \cdot w_N \cdot X_N)]$$

All the moments become for the joint and marginal variables can then be computed based on

$$M_{X,Y}{}^{(l,n)}(0,0) \equiv \frac{\partial^{l+n}}{\partial t_1{}^l \cdot \partial t_2{}^n} M_{X,Y}(t_1,t_2)\bigg|_{t_1=t_2=0} = m_{l,n}$$

or

$$\Phi_{X,Y}{}^{(l,n)}(0,0) \equiv \frac{\partial^{l+n}}{\partial w_1{}^l \cdot \partial w_2{}^n} \Phi_{X,Y}(w_1,w_2)\bigg|_{w_1=w_2=0}$$

and

$$(-j)^{l+n} \cdot \Phi_{X,Y}{}^{(l,n)}(0,0) = m_{l,n}$$

The Central Limit Theorem

The central limit theorem states that the normalized sum of a large number of mutually independent R.V. with zero mean and finite variance tends to the Gaussian normal CDF, provided that the individual variances are small compared to the sum of the variances.

Convolution in "time domain" is multiplication in the "frequency" or "Laplace" domains

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.

Recap of Expected Values

General concept of an expected value

In general, the expected value of a function is:

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) \cdot f_X(x) \cdot dx$$

$$E[g(X)] = \sum_{x=-\infty}^{\infty} g(X) \cdot f_X(x) = \sum_{x=-\infty}^{\infty} g(X) \cdot \Pr(X = x)$$

Means and Variances of Defined density functions.

Practice … calculate the means, 2nd moments and variances for the following:

$$\sigma^2 = E[X^2] - E[X]^2$$

Table 4.3-1 Means, Variances and Mean-Square values for Common Continuous RVs

Family	pdf $f(x)$	Mean $\mu = E[X]$	Variance σ^2	Mean square $E[X^2]$
Uniform	$U(a,b)$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	$\frac{1}{3}(b^2+ab+a^2)$
Exponential	$\frac{1}{\mu}e^{-x/\mu}u(x)$	μ	μ^2	$2\mu^2$
Gaussian	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$\mu^2+\sigma^2$
Laplacian	$\frac{1}{\sqrt{2}\sigma}e^{-\frac{\sqrt{2}}{\sigma}\lvert x\rvert}$	0	σ^2	σ^2
Rayleigh	$\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}u(x)$	$\sqrt{\frac{\pi}{2}}\sigma$	$\left(2-\frac{\pi}{2}\right)\sigma^2$	$2\sigma^2$

Table 4.3-2 Means, Variances, and Mean-Square Values for Common Discrete RVs

Family	PMF $P(k)$	Mean $\mu = E[K]$	Variance σ^2	Mean square $E[K^2]$
Bernoulli	$P_B(k) = \begin{cases} 1, & p \\ 0, & q \end{cases}$ $q \triangleq 1-p$	p	pq	p
Binomial	$b(k;n,p) = \binom{n}{k}p^k q^{n-k}$	np	npq	$(np)^2+npq$
Geometric†	$\frac{1}{1+\mu}\left(\frac{\mu}{1+\mu}\right)^k u(k)$	μ	$\mu+\mu^2$	$\mu+2\mu^2$
Poisson	$\frac{\alpha^k}{k!}e^{-\alpha}u(k)$	α	α	$\alpha^2+\alpha$

Notes and figures are based on or taken from materials in the course textbook: Probability, Statistics and Random Processes for Engineers, 4th ed., Henry Stark and John W. Woods, Pearson Education, Inc., 2012.