Chapter 3 Mathematical Expectation

- Goldberger, Ch. 3
- Wackerly et al. chapter 3-3, 4-3

  Yale note, chapter 5

- Expectation
- Expectation of Function of Random Variables
- Moments
- Expectation and Probability
- Moment Generating Function

1. Expectation of a Random Variable

◎ Expectation: probability weighted average of all possible values the random variable can take.

- Measure of the center of the distribution of X.

(Example) Rolling a dice, $E(X) = 3.5$

- Discrete case: $E(X) = \sum_x xf(x)$,
- Continuous case: $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

(Example) $X \sim U(a,b)$.

2. Expectation of Function of Random Variable

► $X \sim f(x)$

- Discrete case: $E\left[g(X)\right] = \sum_{x} g(x) f(x)$

- Continuous case: $E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x) dx$

(Example) $X \sim U(a,b)$

☞ $E(X^2)$

(Note) $E\left[g(X)\right] \neq g\left[E(X)\right] = g(\mu)$, if g(·) is nonlinear.

(Example) $X \sim U(a,b)$, $E\left[X^2\right] \neq \left[E(X)\right]^2$.

► However, $M\left[g(X)\right] \neq g\left[M(X)\right]$ if $g(\bullet)$ is monotone function.

⊙ Some Important Special Case

- If $g(\cdot)$ is linear, $E\left[g(X)\right]=g\left[E(X)\right]=g(\mu)$.

- Let $a,\ b$ be constants.

① $E(b)=b$

② $E(aX)=a\mu$

③ $E(aX+b)=a\mu+b$

④ $E\left[\sum_{i=1}^{n} a_i g_i(X)\right]=\sum_{i=1}^{n} a_i g_i\left[E(X)\right]$

◎ Moments of Nonlinear Functions of Random Variables(삭제)

Consider a nonlinear function of a random variable X, denoted by $Y = g(X)$.

► Y can be approximated by a linear Taylor series expansion:

$$g(X) \cong g(\mu) + g'(\mu)(X - \mu)$$

Then,

$$E\left[g(X)\right] \cong g(\mu)$$

$$V\left[g(X)\right] \cong g'(\mu)V(X) = \sigma^2 g'(\mu)^2$$

► Will be discussed more in chapter 8(Large Sample Theory).

3. Moments

- *<u>Moments</u>* can be used to describe the shape of the distribution of a random variable.

- Two types of moments: central($(X-\mu)^r$) and uncentral(X^r).

(Definition) The rth *uncentral moment* (*moment about origin*, *raw moment*) of a random variable X is

$$\mu_r' \equiv E\left[X^r\right] = \begin{cases} \sum_x x^r f(x), & \textbf{discrete} \\ \int_{-\infty}^{\infty} x^r f(x)dx, & \textbf{continuous} \end{cases} \qquad r = 0,1,2,\cdots$$

▶ $\mu_0' = 1,\ \ \mu_1' = \mu$.

(Definition) The *rth* <u>*central moment*</u> *(<u>moment about the mean</u>)* of a random variable X is

$$\mu_r \equiv E\left[(X-\mu)^r\right] = \begin{cases} \sum_x (x-\mu)^r f(x), & \textbf{discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^r f(x)dx, & \textbf{continuous} \end{cases} \qquad r = 0,1,2,\cdots.$$

▶ $\mu_0 = 1,\ \mu_1 = 0$.

(Note) $\mu_2 = \sigma^2 = Var(X) = V(X)$,

σ: standard deviation

⊙ Other Important Moments

① $\mu_3 = E\left[(X-\mu)^3\right]$: measure of symmetry

• *Skewness*: $sk = \dfrac{\mu_3}{(\mu_2)^{3/2}} = \dfrac{E\left[(X-\mu)^3\right]}{\left\{E\left[(X-\mu)^2\right]\right\}^{3/2}} = \dfrac{E\left[(X-\mu)^3\right]}{\sigma^3}$

ⓐ When distribution is symmetric ($\mu_3 = 0$), skewness=0.

ⓑ If $sk > 0$, skewed to right (skewed positively).

ⓒ If $sk < 0$, skewed to left (skewed negatively).

ⓓ Unit-free measure.

② $\mu_4 = E\left[(X-\mu)^4\right]$: measure of peakness(flatness)

• *Kurtosis*: $kr = \dfrac{\mu_4}{\sigma^4} = \dfrac{E\left[(X-\mu)^4\right]}{\left\{E\left[(X-\mu)^2\right]\right\}^2}$

► Degree of excess=kr-3

ⓐ kr=3: mesokurtic(중첨) for standard normal distribution.

ⓑ kr>3: leptokurtic(급첨) for sharp-peaked distribution.

ⓒ kr<3: platykurtic(완첨) for flat-peaked distribution.

(Application) stock returns

(Note) $Y = a + bX$ (a, b ; constants) has the same value of skewness & kurtosis as X .

► Other unit free measure?

◎ Theorems on Expectations

① (linear function)

If $Z = a + bX$ where a, b are constants, $E[Z] = a + bE[X]$ and $V[Z] = b^2 V[X]$.

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

③ (<u>mean squared error</u>)

Let c be any constant. Then mean squared error of a random variable about c is $E\left[(X-c)^2\right]=\sigma^2+(c-\mu)^2$

(Note) If $c=E(X)=\mu$, $E\left[(X-c)^2\right]=\sigma^2$

(Note) If c is an estimator of μ, $E\left[(X-c)^2\right]=V\left[X\right]+Bias^2$.

④ (<u>Minimum mean squared error</u>)

The value of c which minimizes $E\left[(X-c)^2\right]$ is $c=\mu$.

► Analogue to sample

◎ Prediction 1(location measure)

- Suppose $X \sim f(x)$, with $f(x)$ known.

A single draw will be made from $f(x)$. You are asked to forecast (predict, guess) the outcome, using a constant c as the predictor.

What is best guess or best predictor?

(Example) 서강대 학생의 수능점수는 몇 점인가?

(Example) 우리 나라 가계의 소득은 얼마인가?

(Theorem) Suppose that your criterion for good predictor is minimum mean squared error. That is, you will choose c to minimize $E\left[U^2\right] = E\left[\left(X - c\right)^2\right]$.

Then, $c^* = \mu$.

- There are many unbiased predictors but μ <u>uniquely</u> minimizes mean squared prediction error.

- For $c^* = \mu$, the forecast error $\varepsilon = X - \mu$, with $E(\varepsilon) = 0$ (⇒unbiased predictor), and

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

- <u>Different criterion for good predictor</u> ⇒ Different choice of predictor

① Minimizing $E(U) = E\left(\left|X - c\right|\right) \Rightarrow c^* = Median(X)$.

② Maximizing $P(U = 0) = P(X = c) \Rightarrow c^* = Mode(X)$.

◎ Expectations of Mixed Distribution

Let Y have the mixed distribution

$$F(y) = c_1 F_1(y) + c_2 F_2(y)$$

and suppose that X_1 is a discrete random variable with distribution function $F_1(y)$ and that X_2 is a continuous random variable with distribution function $F_2(y)$.

Let $g(Y)$ denote a function of Y.

Then

$$E[g(Y)] = c_1 E[g(X_1)] + c_2 E[g(X_2)].$$

(Example) Let Y denote the length of life of electronic components. These components frequently fail immediately with observed probability 1/4. If it does not fail immediately, the distribution for its length of life has exponential density function, $f(y) = e^{-y}, \quad \text{for} \quad y > 0$.

Since $E(X_1) = 0, \quad E(X_2) = 1, \quad E(X_1{}^2) = 0 \quad \text{and} \quad E(X_2{}^2) = 2$.

Then

$$E(Y) = (1/4)0 + (3/4)1 = 3/4.$$

$$E(Y^2) = (1/4)0 + (3/4)2 = 3/2.$$

$$V(Y) = 15/16.$$

4. Expectation and Probability

- Any probability can be interpreted as an expectation:

  Define random variable $Z = \begin{cases} 1 & \text{if event A occurs} \\ 0 & \text{otherwise} \end{cases}$.

  Then, $P(A) = E(Z)$.

- Expectations, variance $\xrightarrow{\text{information}}$ Probability distribution of r. v.

① <u>Markov inequality</u>

If Y is nonnegative random variable ($\Pr(Y<0)=0$) and k is any positive number,

then $P(Y \geq k) \leq \dfrac{E(Y)}{k}$.

② Chebyshev's inequality 1

If X is a random variable, c is a constant and d is any positive constant, then

$$\Pr\left(\left|X-c\right|\geq d\right)\leq\frac{E\left[(X-c)^2\right]}{d^2}.$$

(Proof)

③ <u>Chebyshev's inequality 2</u>

If X is a random variable with $E(X)=\mu$, $V(X)=\sigma^2$, and d is any positive constant, then $\Pr\left(\left|X-\mu\right|\geq d\right)\leq\frac{\sigma^2}{d^2}$.

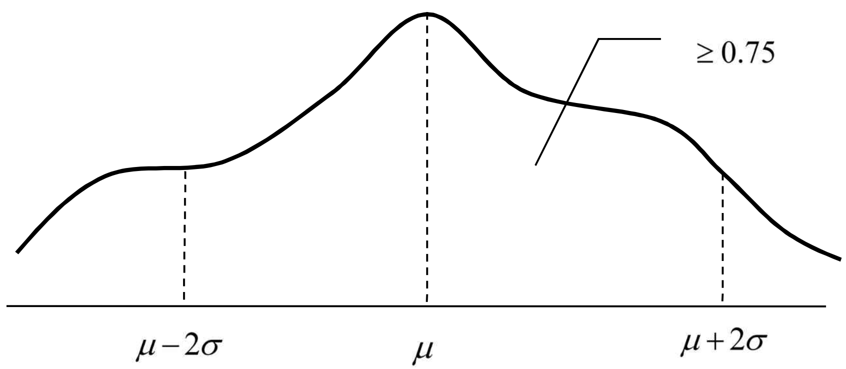
④ <u>Chebyshev's inequality 3</u>

When $d=k\sigma$, $\Pr\left(\left|X-\mu\right|\geq k\sigma\right)\leq\frac{1}{k^2}$ or $\Pr\left(\left|X-\mu\right|<k\sigma\right)\geq 1-\frac{1}{k^2}$.

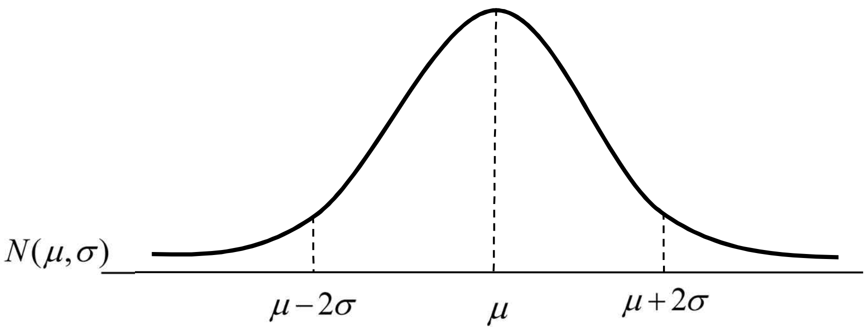
(proof)

(Example)

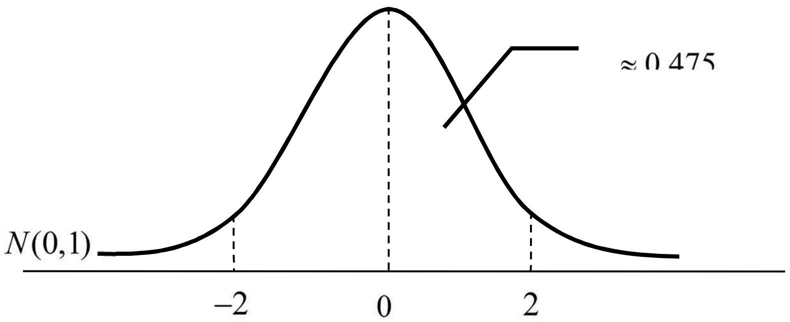
(a) $k = 2 \Rightarrow P\left(\left|X - \mu\right| \le 2\sigma\right) \ge \dfrac{3}{4}$.



(b) $X \sim N(\mu, 1) \;\Rightarrow\; \mathrm{P}(|X-\mu| \le 2) \ge 0.95.$



By Chebyshev's inequality 2,



④ <u>Jensen's inequality</u>

If $Y = h(X)$ is concave and $E(X) = \mu$, then $E\left[Y\right] \le h(\mu) = h\left[E(X)\right]$.

(Cf) If $\mathrm{h}(\cdot)$ is linear, $E(Y) = E\big(h(X)\big) = h\big(E(X)\big) = h(\mu)$.

If $\mathrm{h}(\cdot)$ is nonlinear, $E(Y) = E\big(h(X)\big) \ne h\big(E(X)\big) = h(\mu)$.

(Example) $Var(X) = E(X^2) - \left[E(X)\right]^2 \ge 0$.

5. Moment Generating Function

(Definition)

$$M(t) = E(e^{tX}) = \sum_x e^{tx} f(x) \qquad \text{for discrete case}$$

$$= \int_{-\infty}^{\infty} e^{tX} f(x) dx \qquad \text{for continous case}$$

- Why is it called m.g.f?

$$M'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tX} f(x) dx = \int_{-\infty}^{\infty} x e^{tX} f(x) dx \quad \text{and}$$

$$M'(0) = \int_{-\infty}^{\infty} x f(x) dx = E(X)$$

Furthermore, $M^{(r)}(0) = E(X^r)$

It is called m.g.f. because it generates all the moments of X:

$$\mathrm{E}(X^r) = \frac{d^r}{dt^r} M(t)\Big|_{t=0}.$$

(Example) Poisson Distribution, $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$.

$$M(t) = \exp(\lambda(\mathrm{e}^t - 1))$$

(Example) X~Bernoulli(p). $M(t) = p\,\mathrm{e}^t + (1 - p)$.

(Example) Z~N(0, 1), $M(t) = \exp(t^2 / 2)$.

(Example) $X \sim N(\mu, \sigma^2)$, $M(t) = \exp\left(\mu\,\mathrm{t} + \dfrac{\sigma^2 t^2}{2}\right)$.

◎ Why *m.g.f.*?

• The mgf is unique and completely determines the distribution of the r.v.

• It can be thought of as the DNA of the pdf.

► Once we know the mgf, we know about the pdf.

• Sometimes, it is easier to derive particular moments using the mgf.

• When we are interested in the pdf of functions of r.v., often the only feasible way of deriving the pdf is using the techniques based on the mgf.