Some Notes on the random variables:

- ✓ A random variable, named X, can be viewed as a function from the sample space S to a certain subset of R (the set of real values), denote by $\mathbb{R}_X$. As a function, it satisfies all the properties of functions. In particular,
  - ★ Every element in S has one and only one projected value in $\mathbb{R}_X$.
  - ★ However, for every value in $\mathbb{R}_X$, there may exist an arbitrary number of values in S that may be projected to this value.
- ✓ Such a function X defines "equivalent event" between S and $\mathbb{R}_X$. That is, for any subset B of $\mathbb{R}_X$, there is a subset A of S, such that $B = X(A)$. This idea can be written more mathematically:

$$A = \{s \in S | X(s) \in B\}.$$

  Note that since $\mathbb{R}_X$ is the range for X, we have

$$S = \{s \in S | X(s) \in \mathbb{R}_X\}.$$

- ✓ The goal of introducing the equivalent event is to impose probabilities for the elements (subsets) of $\mathbb{R}_X$, which form the "distribution" of X:

$$Pr(B) = Pr(A).$$

Read the notes above together with the example on page 2-6 of the lecture slides. See also pages 2-13 to 2-16 for a full view of this example.

Example 1

- Let $S = \{HH, HT, TH, TT\}$ be a sample space associated with the experiment of tossing two coins.
- Define the random variable (a function)

  X = number of heads obtained.

  $X : S \to \mathbb{R}$, where $\mathbb{R}$ is the set of all real numbers

  such that $X(HH) = 2, X(HT) = 1, X(TH) = 1$ and $X(TT) = 0$.
- In fact the range space, $\boldsymbol{R_X}$, for the random variable X is

$$\{0, 1, 2\}.$$

2.2 Discrete Probability Distributions

2.2.1 Discrete Random Variable

Definition 2.3

- Let X be a random variable.
- If the number of possible values of X (i.e., $\boldsymbol{R_X}$, the range space) is **finite or countable infinite**, we call X a **discrete** random variable.
- That is, the possible values of X may be listed as $x_1, x_2, x_3, \cdots$.

✓ Countable or uncountable are both concepts for the number of elements for a set with INFINITELY many elements.

✓ A set is called countable, if we can use a "way" to "count" the set; under such a "way", for an arbitrary element in this set, we can clearly speak out when this element will be counted. For example, the set of positive integers is countable, as we can naturally count $1, 2, 3, 4, \ldots$; the set of integers is also countable, as we can count: $0, 1, -1, 2, -2, 3, -3, \ldots$; the set of the real numbers is uncountable; the set of points between 0 and 1 is uncountable.

Note: the mathematical reasoning of why these sets are uncountable belongs to the content of "measure theory", and is beyond the scope of ST2334.

✓ The elements in a countable set can always be listed as: $x_1, x_2, x_3, \ldots$.

Probability Function (Continued)

The probability of $X = x_i$ denoted by $f(x_i)$ (i.e. $f(x_i) = \Pr(X = x_i)$, must satisfy the following two conditions.

(1) $f(x_i) \geq 0$ for all x_i.

(2) $\sum_{i=1}^{\infty} f(x_i) = 1$.

These conditions are induced from the properties of probabilities defined on the sample space. In particular, if we define $A_i = \{s \in S | X(s) = x_i\}$, then based on the concepts we discussed in the first page of this document, we have:

★ A_i is the "equivalent set" of $\{x_i\}$, therefore

$$f(x_i) = Pr(X = x_i) = Pr(A_i).$$

★ As $X(\cdot)$ is a "function", therefore, A_i's are disjoint sets.

★ $\mathbb{R}_X$ is the range for X, so $\cup_{i=1}^{\infty} A_i = S$.

Therefore, $1 = Pr(S) = \sum_{i=1}^{\infty} Pr(A_i) = \sum_{i=1}^{\infty} f(x_i)$.

When establishing the probability distribution in some practical problems, make sure that this criteria is satisfied!

Page 2-35

Example 5

- Consider a group of five potential blood donors — A, B, C, D and E — of whom only A and B have type O+ blood.
- Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified.

Prof. Chan used the counting method to solve the question in the lecture video; the lecture slides used the conditional probability method. Make sure that you are able to solve this question by either of these methods.

2.3 Continuous Probability Distributions

2.3.1 Continuous Random Variable

Definition 2.4

- Suppose that R_X, the range space of a random variable, X, is an **interval or a collection of intervals.**
- Then we say that X is a **continuous random variable.**

✓ For a discrete random variable, its range must be finite or countable; the random variable has a point mass on each possible value in its range. Here "has a point mass" typically refers to it has positive probability to take the value; for example, "X has a point mass on x_i" means "$P(X = x_i) > 0$".

✓ For a continuous random variable, its range is an interval or a collection of intervals, which means its range is not countable. It has NO point mass on any particular value in its range. This means that for any $x \in \mathbb{R}_X$, we must have $Pr(X = x) = 0$.

This immediately leads to a very good example for the statement: "for a set $B \subset \mathbb{R}_X$, $Pr(X \in B) = 0$ does not imply $B = \emptyset$. " In fact, for any $B \subset \mathbb{R}_X$, if the number of elements in B is countable, we conclude $Pr(X \in B) = 0$.

This also implies if X is a continuous random variable, it is impossible for X to take any particular value. For example, height is a continuous variable and it is impossible to know the exact height of a subject. The value that we get is up to the accuracy provided by an instrument that we use.

✓ Do we have random variables that are in between? Say, can we find a random variable whose range is a collection of intervals, but has point mass on some values in its range? The answer is yes; but is not the focus of this module. Please try to find one such random variable on your own.

Page 2-43 & 44

2.3.2 Probability Density Function

Definition 2.5

- Let X be a continuous random variable.
- The **probability density function (p.d.f.)** $\boldsymbol{f(x)}$, is a function, $f(x)$, satisfying the following conditions:
  1. $f(x) \geq 0$ for all $x \in \boldsymbol{R_X}$,
  2. $\int_{\boldsymbol{R_X}} f(x)\,dx = 1$ or $\int_{-\infty}^{\infty} f(x)\,dx = 1$

     since $f(x) = 0$ for x not in $\boldsymbol{R_X}$.

ST2334 Probability and Statistics CYM Concepts of Random Variables 2-43

Probability Density Function (Continued)

Definition 2.5 (Continued)

3. For any c and d such that $c < d$, (i.e. $(c, d) \subset \boldsymbol{R_X}$),

$$\Pr(c \leq X \leq d) = \int_c^d f(x)dx$$

ST2334 Probability and Statistics CYM Concepts of Random Variables 2-44

The probability density function (p.d.f.) plays a very similar role as the probability mass function (p.m.f.) for the discrete case. But one should keep in mind that they also have some major difference:

- ✓ If $f(x)$ is a p.m.f., we must have $f(x) \leq 1$ for any x. In contrast, if $f(x)$ is a p.d.f., this is not necessarily true.

- ✓ The value of a p.m.f. $f(x)$ has a very obvious probability meaning: it is the probability that the random variable will take the value x. But the value of a p.d.f. $f(x)$ does not have such a meaning (note that it makes nonsense to talk about the probability that the random variable will take a specific value, as this probability is always 0); in stead, the probability that the random variable will be in an interval is evaluated by the areas under the function curve of $f(x)$; the corresponding mathematical statement has been given by item 3 of Definition 2.5 on page 2-44 of the lecture slides.

  The p.d.f. reflects how likely the corresponding random variable will fall in a very small neighbourhood of X; namely for a sufficiently small δ, $P(x_0 \leq X \leq x_0 + \delta) \approx f(x_0)\delta$.

For item 2 of Definition 2.5 on page 2-43, we take note that $\int_{\mathbb{R}_X} f(x)dx = 1$ and $\int_{-\infty}^{\infty} f(x)dx = 1$ are exactly the same. This is because for $x \notin \mathbb{R}_X$, $f(x) = 0$, therefore

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{x\in\mathbb{R}_X} f(x)dx + \int_{x\notin\mathbb{R}_X} f(x)dx \\
&= \int_{x\in\mathbb{R}_X} f(x)dx + \int_{x\notin\mathbb{R}_X} 0dx = \int_{x\in\mathbb{R}_X} f(x)dx.
\end{aligned}$$

Remarks (Continued)

2. For any specified value of X, say x_0, we have
$$\Pr(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$$
Hence in the **continuous** case, the probability of X equals to a fixed value is 0 and
$\Pr(c \le X \le d) = \Pr(c \le X < d) = \Pr(c < X \le d) = \Pr(c < X < d)$.
Therefore in the continuous case, $\le$ **and** $<$ can be used **interchangeably** in a probability statement.

Take note of the formulae

$$Pr(c \le X \le d) = Pr(c \le X < d) = Pr(c < X \le d) = Pr(c < X < d),$$

and be aware that they are applicable only when X is a continuous random variable. In general, we have

$$\begin{aligned} Pr(c \le X \le d) &= Pr(c \le X < d) + Pr(X = d) \\ &= Pr(c < X \le d) + Pr(X = c) \\ &= Pr(c < X < d) + Pr(X = c) + Pr(X = d). \end{aligned}$$

Therefore, when X is a discrete random variable, we need to account for whether X has point masses on c and d.

Example 2 (Continued)

- Clearly, $f(x) \geq 0$ and

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0.5} 0\ dx = \int_{0.5}^{\infty} 0.15\ e^{-0.15(x-0.5)}\ dx$$

$$= 0.15e^{0.075} \int_{0.5}^{\infty} e^{-0.15x}\ dx$$

$$= 0.15e^{0.075} \left[-\frac{1}{0.15} e^{-0.15x} \right]_{0.5}^{\infty}$$

$$= 0.15e^{0.075} \left(0 - \left(-\frac{1}{0.15} \mathrm{e}^{-0.15(0.5)} \right) \right) = 1$$

In line with the lecture video, $\int_{-\infty}^{\infty} f(x)dx = 1$ needs to be checked only when the question asks you to check. For some distributions, checking this may not be an easy task!

2.4 Cumulative Distribution Function

Definition 2.6

- Let X be a random variable, discrete or continuous.

- We define $F(x)$ to be the **cumulative distribution function** of the random variable X (abbreviated as c.d.f.) where

$$F(x) = \Pr(X \leq x).$$

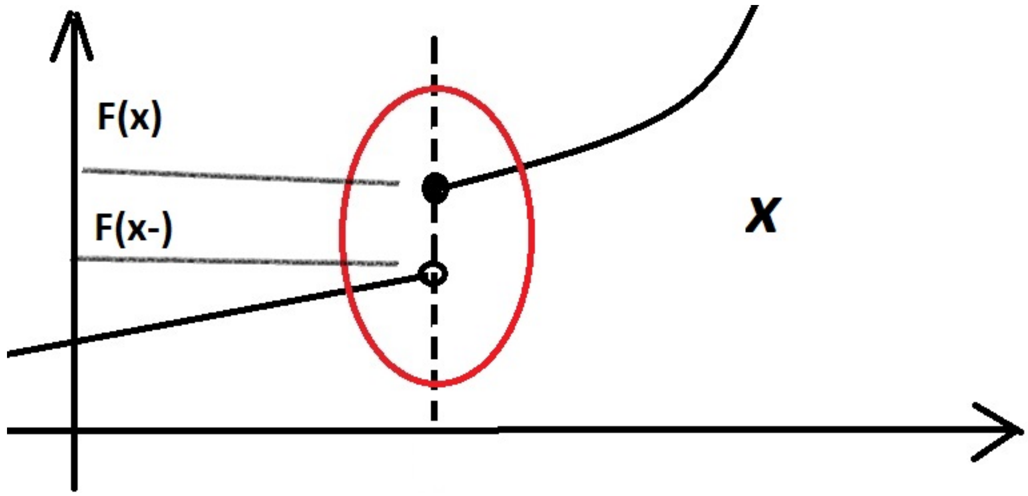
For any random variable, discrete or continuous or "in between", the cumulative distribution function (c.d.f.) is always defined as the one given in this slide.

If $F(x)$ is a c.d.f. for some random variable, then it satisfies

✓ $F(x)$ is a nondecreasing function of x.

✓ We always have $F(x) \to 1$, as $x \to \infty$; $F(x) \to 0$, as $x \to -\infty$.

✓ For every $x \in \mathbb{R}$, $F(x)$ is either continuous, or if it is not continuous, it must be right continuous at x and the left limit exists. In a figure, it must be like this:



Mathematically, it means for any x

$$\lim_{t \to x+} F(t) = F(x) \qquad \lim_{t \to x-} F(t) \text{ exists,}$$

the convergence result of the latter one is denoted as $F(x-)$. See the difference of $F(x)$ and $F(x-)$ in the figure. We have $Pr(X = x) = F(x) - F(x-)$, which is the point mass of random X at x.

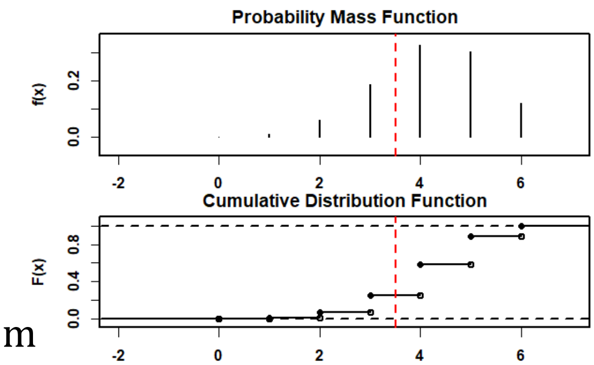
✓ When X is a discrete random variable, its c.d.f. must be a step function in a similar structure as the one given in page 2-60 of the lecture slide. Take note that the mathematical formula of its c.d.f. is also given in this slide.

2.4.1 CDF for Discrete Random Variables

- If X is a **discrete random variable**, then

$$F(x) = \sum_{t \le x} f(t)$$
$$= \sum_{t \le x} \Pr(X = t)$$

- The c.d.f. of a discrete random variable is a step function.



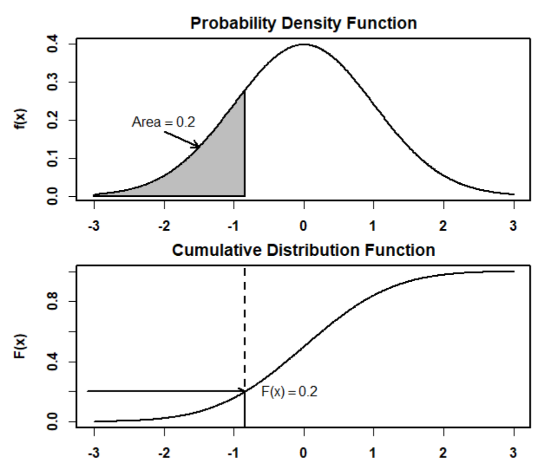
ST2334 Probability and Statistics CYM Concepts of Random Variables 2-60

✓ When X is a continuous random variable, its c.d.f. must be a continuous function, with the c.d.f. in a similar structure as the one given in page 2-63 of the lecture slides. This page also provides the mathematical formula of its c.d.f.

2.4.2 CDF for Continuous Random Variables

- If X is a **continuous random variable**, then

$$F(x) = \int_{-\infty}^{x} f(t)dt$$



ST2334 Probability and Statistics CYM Concepts of Random Variables 2-63

CDF for Continuous Random Variables (Continued)

- For a **continuous random variable** X,
$$f(x) = \frac{d\,F(x)}{dx}$$
if the derivative exists.

- Also,
$$\begin{aligned}\Pr(a \le X \le b) &= \Pr(a < X \le b)\\ &= F(b) - F(a).\end{aligned}$$

This page gives how we can obtain the p.d.f. when we have the c.d.f. of a continuous random variable in hand.

When X is a discrete random variable, we have $f(x) = F(x) - F(x-)$. So,

✓ when $F(x)$ is continuous at x, $f(x) = 0$;

✓ when $F(x)$ is not continuous at x, then $f(x) > 0$;

therefore, the discontinuous points of $F(x)$ attributes to all the point masses of the corresponding p.d.f.

Page 2-81

Solution to Example 5

(a) It is obvious that $f(x) > 0$ for $x > 0$.

$$\int_{-\infty}^{\infty} f(x;\theta)\,dx = \int_0^{\infty} \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}\,dx$$
$$= \int_0^{\infty} d\left(e^{-\frac{x^2}{2\theta^2}}\right)$$
$$= \left[e^{-\frac{x^2}{2\theta^2}}\right]_0^{\infty}$$
$$= 0 - (-1) = 1.$$

A small amendment:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x;\theta)dx &= \int_0^{\infty} \frac{x^2}{\theta} e^{-\frac{x^2}{2\theta^2}} dx \\ &= -\int_0^{\infty} d\left(e^{-\frac{x^2}{2\theta^2}}\right) \\ &= -\left[e^{-\frac{x^2}{2\theta^2}}\right]_0^{\infty} \\ &= -[0-1] = 1.\end{aligned}$$

Page 2-90

Remarks

1. The expected value exists provided the sum or the integral in the above definitions exists.

2. In the discrete case, if $f_X(x) = 1/N$ for each of the N values of x, hence the mean,

$$E(X) = \sum_i x_i\, f(x_i) = \frac{1}{N}\sum_i x_i\,,$$

becomes the average of the N items.

We give one example that the expectation does not exist:

Assume that X takes point masses on $1^2, 2^2, 3^2, \ldots,$; and for each $n = 1, 2, \ldots,$

$$P(X = n^2) = \frac{6}{\pi^2}\frac{1}{n^2}.$$

Note that it is well known $\sum_{n=1}^{\infty} \frac{1}{n^2} = \pi^2/6$, which is called the Basal's problem, and the solution was first found by Euler in 1735; therefore, this gives an appropriately defined (discrete) probability distribution. However, if we try to evaluate $E(X)$ using the formula given in page 2-87, we have

$$E(X) = \sum_{n=1}^{\infty} n^2 \frac{1}{n^2}\frac{6}{\pi^2} = \frac{6}{\pi^2}\sum_{n=1}^{\infty} 1 = \infty.$$

Please try on your own to find an example that the expectation of a continuous random variable does not exist.

Pages 2-91 & 92

Example 1

- In a gambling game, a man gains 5 if he gets all heads or all tails in tossing a fair coin 3 times,
- and he pays out 3 if either 1 or 2 heads show.
- What is his expected gain?

Solution to Example 1

- Let X be the amount he can gain.
- Then $X = 5$ or -3 with the following probabilities:
  $\Pr(X = 5) = \Pr(\{HHH, TTT\}) = 1/8 + 1/8 = 1/4$
  $\Pr(X = -3) = 1 - \Pr(X = 5) = 3/4.$
- Therefore $E(X) = 5\left(\frac{1}{4}\right) + (-3)\left(\frac{3}{4}\right) = -1.$
- Hence, he will lose 1 per toss **in a long run**.

How could we change the pay amount in 2nd item "and he pays 3 if either 1 or 2 heads show" on page 2-91 so that the game is a fair game?

Here fair means that the expected gain would be equal to 0.

We can set the amount he pays out if either 1 or 2 heads show is a. Then the expected gain would be

$$E(X) = 5(1/4) + (-a)(3/4)$$

Then setting this expected value to be 0 and solve the equation for a, we can get $a = 5/3$, which is the amount to replace 3 so that the game is fair.

Certainly 5/3 is not a practical amount to pay in a single game. Is there any other way that we can adjust to make the game fair? Try to adjust the gains when all heads and all tails are obtained in the game!

Pages 2-93 & 94

Example 2

- Suppose a game consists of rolling a balanced die.
- We pay c to play the game and we get i if number i occurs.
- How much should we pay if the game is fair?
  (We say a game is "fair" if $E(\text{gain}) = 0$.)

Solution to Example 2

- Let X denote the amount that one gets when rolling a die.
- Then clearly $\Pr(X = 1) = \cdots = \Pr(X = 6) = 1/6$.
- $E(X) = (1 + 2 + \cdots + 6)\left(\frac{1}{6}\right) = 3.5$.
- To be a fair game, $E(\text{paying}) = E(\text{getting})$, and hence for a fair game, the admission fee should be $c = E(X) = 3.5$.

Please get familiar with the computation of the expectation for discrete random variables; it is the "weighted average" of the possible values of the random variable, where "weights" are the corresponding probabilities.

In this example, if the die is unfair, such that the probability of getting 1 is 0.5, and getting 2, 3,4,5,6 are all 0.1, then

$$E(X) = 0.5 \cdot 1 + 0.1 \cdot (2 + 3 + 4 + 5 + 6) = 2.5.$$

So to ensure that the game is fair, we need to set $c = 2.5$, which is lower than 3.5. This intuitively makes sense. As we have greater chance to get the lowest reward 1, to balance such a results, we need to pay less to play the game so that we won't "expect" to lose in the long run.

2.5.2 Expectation of a Function of a RV

Definition 2.8

For any function $g(X)$ of a random variable X with p.f. (or p.d.f.) $f_X(x)$,

(a) $E[g(X)] = \sum_x g(x) f_X(x)$

if X is a **discrete** r.v. providing the sum exists; and

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

if X is a **continuous** r.v. providing the integral exists.

This page gives very important and useful formulae. These formulae allow us to compute the expectations of an arbitrary function of a random variable with a known distribution.

For example, if we have a continuous random variable X, we know its p.d.f. is $f_X(x)$, and we want to find $E(\sin(X))$, we do not need to derive the p.d.f. for $\sin(X)$, instead, we only need to apply the formula $E(\sin(X)) = \int_{-\infty}^{\infty} \sin(X) f_X(x) dx$ to compute this expectation.

Some discussion on the expectation $E(X)$ and variance $V(X)$:

✓ Expectation is essentially given the "population mean" (or intuitively the "central location" of the possible values) of X. Therefore, one may also see that $E(X)$ is called the location parameter in some literature.

✓ Variance tells how the possible values of X spread around $E(X)$. The smaller the $V(X)$, the more concentrated the possible values of X around $E(X)$; and vice versa. $V(X)$ is frequently called the dispersion parameter in many places.

✓ Note that $V(X)$ has two alternative (but in fact equivalent) formulae to get:

$$V(X) = E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2 . \tag{1}$$

★ The former serves as the definition of variance given on page 2-105.

Some Special Cases (Continued)

Definition 2.9

- Let X be a random variable with p.f. (or p.d.f.) $f(x)$, then the **variance** of X is defined as

$$\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$$

$$= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x), & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx, & \text{if } X \text{ is continuous.} \end{cases}$$

ST2334 Probability and Statistics CYM Concepts of Random Variables 2-105

★ The latter is mathematically derived on page 2-126.

Properties of Expectation (Continued)

Property 2 (Continued)

Proof:

$$V(X) = E[(X-\mu_X)^2]$$
$$= E[X^2 - 2X\mu_X + \mu_X^2]$$
$$= E(X^2) - E(2X\mu_X) + E(\mu_X^2)$$
$$= E(X^2) - 2\mu_X\, E(X) + (\mu_X^2)$$
$$= E(X^2) - 2\mu_X^2 + \mu_X^2 = E(X^2) - \mu_X^2$$

Notice that $\mu_X = E(X)$ is a constant.

★ The latter formula is usually computationally more convenient when we have the p.d.f. available and we are to compute the variance. Examples 1–3 on page 2-108 to 2-121 give good illustrate of this.

✓ One very obvious observation based on (1) is that $E(X^2) \geq (E(X))^2$, and "=" can hold if and only if X is a constant (i.e., nonrandom); more specifically, there is a given value c, such that $P(X=c)=1$. In such a case, $V(X)=0$ (i.e., no variability for X).

✓ Several very useful formulae for expectation and variance are given below:

★ For the expectation, it has the linearity: for ANY random variables $X_1, X_2, \ldots, X_i$, and constants (nonrandom) $a_1, a_2, \ldots, a_k$,

$$E(a_1X_1 + a_2X_2 + \ldots + a_kX_k) = a_1E(X_1) + a_2E(X_2) + \ldots + a_kE(X_k).$$

This is a slightly more general formula than that given on page 2-124. Think about why.

Properties of Expectation (Continued)

Two special cases:

(a) Put $b = 0$, we have $E(aX) = a\,E(X)$.

(b) Put $a = 1$, we have $E(X + b) = E(X) + b$.

In general,

$$E[a_1 g_1(X) + a_2 g_2(X) + \cdots + a_k g_k(X)]$$
$$= a_1 E[g_1(X)] + a_2 E[g_2(X)] + \cdots + a_k E[g_k(X)]$$

where $a_1, a_2, \cdots, a_k$ are constants.

★ For the variance, we have the formula:

$$V(aX + b) = V(aX) = a^2 V(X),$$

which can be viewed as the combination of two formulae:

- for any random variable Y and constant b, $V(Y + b) = V(Y)$;

$$V(Y+b) = E\left(((Y+b) - E(Y+b))\right) = E\left((Y - E(Y))^2\right) = V(Y);$$

- and for any constant a and random variable X, $V(aX) = a^2 V(X)$;

$$V(aX) = E\left((aX - E(aX))^2\right) = a^2 E\left((X - E(X))^2\right) = a^2 V(X).$$

The development above can also be used to replace the proof given on page 2-128 of the lecture slides.

Properties of Expectation (Continued)

Property 3 (Continued)

Proof:

$$V(aX+b) = E[(aX+b)^2] - [E(aX+b)]^2$$
$$= E(a^2X^2 + 2abX + b^2) - (a\mu_X + b)^2$$
$$= a^2E(X^2) + 2ab\,E(X) + b^2 - (a^2\mu_X^2 + 2ab\mu_X + b^2)$$
$$= a^2E(X^2) - a^2\mu_X^2$$
$$= a^2[E(X^2) - \mu_X^2]$$
$$= a^2V(X).$$

✓ In practice, we may jointly consider the expectation and the variance to judge the performance of a random variable. See the example given on page 2-129. The expected profit (700$) derived on page 130 looks promising. The standard deviation (800$ given on page 2-131), however, is much too big. Eventually, the chance that the shop will lose money is not small (for this specific example, you can make scenarios that the shop may lose money; you can also compute the probability that the shop will lose money on your own), especially in a short period of time. So, if you are the boss of the shop, you might think of whether you want to take the risk.

Example 4

- A jewelry shop purchased three necklaces of a certain type at \$500 a piece.
- It will sell them for \$1000 a piece. The designer has agreed to repurchase any necklace still unsold after a specified period at \$200 a piece.
- Let X denote the number of necklaces sold and suppose X follows the following probability distribution.

x	0	1	2	3
$f_X(x)$	0.1	0.2	0.3	0.4

- Find the expected gain and the variance of the gain.

Solution to Example 4

- With $g(X) = \text{revenue} - \text{cost} = 1000X + 200(3 - X) - 3(500) = 800X - 900$.
- $E(g(X)) = g(0)f_X(0) + g(1)f_X(1) + g(2)f_X(2) + g(3)f_X(3)$
  $= (-900)(0.1) + (-100)(0.2) + (700)(0.3) + 1500(0.4)$
  $= 700.$
- Hence the expected profit is \$700.

Solution to Example 4 (Continued)

- $E\left[\left(g(X)\right)^2\right] = (-900)^2(0.1) + (-100)^2(0.2) + (700)^2(0.3) + 1500^2(0.4) = 1130000$.

- Hence

$$V\left(g(X)\right) = 1130000 - 700^2 = 640000$$

and

$$\sqrt{V\left(g(X)\right)} = \sqrt{640000} = 800$$

Chebyshev's Inequality (Continued)

- Let X be a random variable (discrete or continuous) with $E(X) = \mu$ and $V(X) = \sigma^2$.
- Then for any positive number k we have

$$\Pr(|X - \mu| \geq k\sigma) \leq 1/k^2.$$

- That is, the probability that the value of X lies at least k standard deviation from its mean is at most $1/k^2$.
- Alternatively,

$$\Pr(|X - \mu| < k\sigma) \geq 1 - 1/k^2.$$

✓ Make clear that Chebyshev's inequality only provides some bounds for the probabilities that the random variable will take a values in a certain range; these are not sharp bounds in most cases.

On the other hand, it definitely plays fundamental roles in the developments of probability and statistical theories, though these are beyond the scope of this module.

✓ A slightly more handy forms of these formulae are given below

$$\begin{aligned} Pr(|X - \mu| \geq c) &\leq \frac{V(X)}{c^2} \\ Pr(|X - \mu| < c) &\geq 1 - \frac{V(X)}{c^2}, \end{aligned}$$

for any constant $c > 0$. Bear in mind these are exactly the same formulae as those given in the slide.