Lecture 4 : Random variable and expectation

Study Objectives: to learn the concept of

1. Random variable (rv), including discrete rv and continuous rv; and the distribution functions (pmf, pdf and cdf).
2. Conditional expectation and conditional distribution; independence.
3. Expectation, variance and covariance.
4. Weak law of large numbers.
5. Moment generating function.

Random Variable: examples

- Example 1:

X= the outcome of a fair dice ➔
We call X a discrete 'random variable'.

-- Pr(X=1)=1/6, …., Pr(X=6)=1/6;

-- $\Sigma_{\{x\}}$Pr(X=***x***)=1, {***x***}={1,2,3,4,5,6}

- Example 1 (cont.) See textbook page 89~90 for a **joint distribution** of X=X1+X2, X1 and X2 are the outcomes of two independent fair dices.

- **Example**: Two electronic components with four possible results: Pr(d,d)=0.09, Pr(d,a)=0.21, Pr(a,d)=0.21, Pr(a,a)=0.49; where d and a denote *defective* and *acceptable* respectively.

(1) If X=number of acceptable components

$\Rightarrow$ Pr(X=0)=0.09, Pr(X=1)=0.42, Pr(X=2)=0.49

(2) If we further define $I = \begin{cases} 1, & \text{if } X=1 \text{ or } 2 \\ 0, & \text{if } X=0 \end{cases}$

$\Rightarrow$ Pr(I=1)=0.91

Pr(I=0)=0.09

Note:

If **A** denotes the event that at least one acceptable component is obtained, the random variable ***I*** is called the **indicator** random variable for the (occurrence of) event **A**, because ***I*** will equal 1 or 0 depending on whether **A** occurs.

Types of random variables

- Discrete or categorical

If the possible values of X is a sequence.

Probability mass function:

p(a)=P(X=a); p(a) is positive for at most a countable numbers of values of a.=> see **Example 4.2a** (textbook, page 92.)

- Continuous

The possible values of X is an interval and is a subset of $(-\infty,\infty)$

Cumulative distribution function (cdf)

$$F(a)=\sum_{\{\boldsymbol{x}:\ \boldsymbol{x}<=a\}} p(\boldsymbol{x})$$

- $F(x)$, or $F_{\mathbf{X}}(x)$,$=\Pr(\mathbf{X}\leq x)$, $\mathbf{X}\sim F(\bullet)$

Read: X be distributed as F

Property:

$P(a<\mathbf{X}\leq b)=P(\mathbf{X}\leq b)-P(\mathbf{X}\leq a)$

$= F(b)-F(a)$

- Example: $F(x)=0$ for $x \leq 0$ and $F(x)=1-\exp(-x^2)$ for $x>0$=> $P(\mathbf{X}>1)=1-P(\mathbf{X}\leq 1)=1-\{1-\exp(-1)\}$

$= \exp(-1)=0.368$

Probability distribution function (pdf)

- Continuous random variable
- f(x) is called a pdf if
  - $\int_S f(x)=1$, where S is the 'support' of the random variable X, or the distribution f(•), indicating the 'range' where f(x)>0.
  - Let $I=[a,b] \subset S$, => $\Pr(X \in I)= \int_I f(x)dx$
  - $dF(x)/dx|_{x=a}=f(a)$, and $F(a)= \int_{\{I=(-\infty,a)\}} f(x)dx$
- Example: [textbook page 95, **example 4.2b**]

Joint distribution of two random variables (continuous case)

X and Y are both continuous rv with support $(-\infty,\infty)$

- Joint cdf of X and Y

  $F(x,y)=P\{X \leq x, Y \leq y\}$

  $= \int_{(-\infty,x)} \int_{(-\infty,y)} f_X(x)f_Y(y)dxdy$

- Of course, $\int_{(-\infty,\infty)} \int_{(-\infty,\infty)} f_X(x)f_Y(y)dxdy=1$
- The distribution of X can be obtained as:

  $F_X(x)=P\{X \leq x\}= P\{X \leq x, Y< \infty\}=F(x, \infty)$

  $= \int_{(-\infty,x)} \int_{(-\infty, \infty)} f_X(x)f_Y(y)dxdy$

Joint distribution of two random variables (discrete case)

- $p(x_i,y_j)=P\{X=x_i,Y=y_j\}$
- $P(X=x_i)=P(\cup_{\mathbf{j}}\{X=x_i,Y=y_j\})$

  $=\sum_{\mathbf{j}}P\{X=x_i,Y=y_j\}$

  $=\sum_{\mathbf{j}}p(x_i,y_j)$
- Example: [See textbook, page 97~98, **examples 4.3a,b,c**]

Independent random variables

- **Continuous case:** **X** and **Y** are said to be independent if $P\{\mathbf{X}\leqq a,\mathbf{Y}\leqq b\}=P\{\mathbf{X}\leqq a\}P\{\mathbf{Y}\leqq b\}$, or

  $F_{\mathbf{X},\mathbf{Y}}(a,b)=F_{\mathbf{X}}(a)\,F_{\mathbf{Y}}(b)$, for all a and b;

  or $f_{\mathbf{X},\mathbf{Y}}(a,b)=f_{\mathbf{X}}(a)\,f_{\mathbf{Y}}(b)$,
- **Discrete case:** **X** and **Y** are said to be independent

  if $p_{\mathbf{X},\mathbf{Y}}(x,y)=p_{\mathbf{X}}(x)p_{\mathbf{Y}}(y)$
- **Example**: [Textbook page 102~104, **examples 4.3d ,e**]

Conditional distributions

- $p_{X|Y}(x|y)=P\{X=x|Y=y\}$ Conditional probability based on events

  probability mass $=P\{X=x,Y=y\}/P\{Y=y\}$

  $=p_{X,Y}(x,y)/p_Y(y)$ Discrete case

- $f_{X|Y}(x|y)dx=f_{X,Y}(x,y)dxdy/f_Y(y)dy$ Continuous case

  $\Rightarrow f_{X|Y}(x|y)=f_{X,Y}(x,y)/f_Y(y)$

  note: $\int f_{X|Y}(x|y)f_Y(y)dy=\int f_{X,Y}(x,y)dy$

  $=f_X(x)$

- **Example**: [Textbook page 105~107, **examples 4.3f ,g,h**]

Expectation

- E(**X**), or simply E**X**, and is common denoted as μ, is defined as

  $E\mathbf{X}=\sum_{\{x\}} xP\{\mathbf{X}=x\}$, for discrete rv

  $=\int_{\{x\}} xf_X(x)dx$, for continuous rv

- **Properties**:

  $E[g(x)]=\int_{\{x\}} g(x)f_X(x)dx$

  $E(a\mathbf{X})=aE\mathbf{X}$, for any constant ‘a’.

  $E(\mathbf{X}+\mathbf{Y})=E\mathbf{X}+E\mathbf{Y}$

- **Example**: [Textbook page 112~118, **examples 4.5a~h**]

Variance

- $\mathrm{Var}(\mathbf{X})=E\{(\mathbf{X}-\mu)^2\}$, $\mu=E(\mathbf{X})$.
- **Property 1**: $\mathrm{Var}(\mathbf{X})=E(\mathbf{X}^2)-\mu^2$

  $[\Rightarrow \mathrm{Var}(a\mathbf{X})=a^2\mathrm{Var}(\mathbf{X})]$

For the case of two rv's **X** and **Y** with some joint distribution, we have:

- **Property 2**: $E\{E[X|Y]\}=EX$; and,

  $E\{E[g(X)|Y]\}=E\{g(X)\}$

  Exercise
- **Property 3**: $\mathrm{Var}(X)=E[\mathrm{Var}(X|Y)]+\mathrm{Var}[E(X|Y)]$

Justification of Property 3

- 1. $E\{\mathrm{Var}(X|Y)\}=E\{E(X^2|Y)\}-E\{[g(Y)]^2\}$ by definition, where $g(Y)=E(X|Y)$;
- 2. $\mathrm{Var}\{E(X|Y)\}= E\{[g(Y)]^2\}-E(X^2)$ can be easily obtained;
- Combining 1 and 2 we have

  $E\{\mathrm{Var}(X|Y)\}+\mathrm{Var}\{E(X|Y)\}= E\{E(X^2|Y)\}-E(X^2)$,

  which is simply $\mathrm{Var}(X)$ using the **Property 2**.

Covariance and correlation

- The covariance of two rv's X and Y is defined as: cov(X,Y)=E{(X-EX)(Y-EY)}, which can be easily re-expressed as E(XY)-(EX)(EY)
- Var(X+Y)=Var(X)+2cov(X,Y)+Var(Y) and
  Var(X-Y)=Var(X)-2cov(X,Y)+Var(Y)
- cov(aX,Y)=acov(X,Y) and cov(X,aY)=acov(X,Y)
- cov(X+Z,Y)=cov(X,Y)+cov(Z,Y)
- Propositions 4.7.2 and 4.7.3
- corr(X,Y)=cov(X,Y)/$\sqrt{}$(Var(X)) $\sqrt{}$(Var(Y))

Moment generating functions

- The moment generating function (mgf) of the random variable **X** is defined as g(t)=E{exp(t**X**)}.
- For discrete case, g(t)= $\sum_{\{x\}}$exp(t**X**)p(***x***)
- For continuous case, g(t)= $\int_{\{x\}}$exp(t**X**)f(***x***)d***x***
- Let $g^{(n)}(t)=d^n g(t)/dt^n$, for n>=1; then we have:
  $g^{(n)}(0)=E(\mathbf{X}^n)$, explains why g is called a mgf, and,
  $g_{\mathbf{X+Y}}(t)=g_{\mathbf{X}}(t)g_{\mathbf{Y}}(t)$
- **Theorem**: The mgf uniquely determines the dsitribution; which means: a specific mgf corresponds to a unique distribution, and *vice versa*.

The weak law of large numbers (WLLN)

- Let $X_1,\ldots\ldots,X_n$ be a set of iid rv's with mean μ. Then for any $\varepsilon > 0$,

  $P\{|(X_1+\ldots+X_n)/n-\mu| > \varepsilon\}$ converges to 0.
- By Chebyshev's inequality, it can be proved that

  $P\{|(X_1+\ldots+X_n)/n-\mu| > \varepsilon\} < Var\{(X_1+\ldots+X_n)/n\}/\varepsilon^2=(\sigma^2/n)\varepsilon^2$. It certainly converges to zero.
- We say: $(X_1+\ldots+X_n)/n \quad \mu$ in probability (*in pr*); in other words, the 'sample mean' is a *consistent* estimator of the parameter μ, the population mean.

Homework and exercises

1. State and prove the Markov inequality and the Chebyshev's inequality.
2. Please complete the proof of Property 3:

   $$Var(X)=E[Var(X|Y)]+Var[E(X|Y)]$$
3. Let $T(x_1,\ldots x_n)$ be unbiased for estimating μ and $Var\{T(x_1,\ldots x_n)\}$ converges to zero when $n \uparrow \infty$. Please show that the T-statistic is consistent for μ.

Homework and exercises *(cont.)*

4. Do the following problems in your textbook pages 130~140:

[**Level 1**]: 4, 6, 8, 12, 30, 36, 45, 49, 52

[**Level 2**]: 10, 11, 16, 28, 29, 35, 40, 50, 54, 57