Ch2) 1) A ——— 13

is a real valued function for which the domain is a sample space.

ex) flip fair coin twice $S = \{HH, HT, TH, TT\}$.

Let Y = # of heads. $\underbrace{Y(HH)=2,\ Y(HT)=Y(TH)=1,\ Y(TT)=0.}_{\text{usually do not use this notation}}$

y	0	1	2
$P(Y=y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

} use this notation

2) The event $\{Y=y\} = \{E_i \in S : Y(E_i) = y\}$.

Hence $P(Y=y)$ is the sum of all sample points in S that are assigned the value y.

ex) p23 $I_E = \begin{cases} 1 & \text{event } E \text{ occurs} \\ 0 & \text{event } E \text{ does not occur} \end{cases}$

I_E is the indicator RV for event E.

p24

3) know for final! The cumulative distribution function (cdf) of a RV X is $F(x) = P(X \le x)$ for $x \in (-\infty, \infty) = \mathbb{R}$.

4) If $F(x)$ is a cdf, then

df1) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$

df2) $F(\infty) = \lim_{x \to \infty} F(x) = 1$

df3) $F(x)$ is a nondecreasing function of x! if $x_1 < x_2$, then $F(x_1) \le F(x_2)$.

not in book) d4) $F(x)$ is right continuous $F(x) = F(x_0)$ for any x_0, $x \downarrow x_0$

13.5

Note: if $A = \{X < x_1\}$ and $B = \{X < x_2\}$ and $x_1 < x_2$, then $A \subseteq B$. Also $\{Y \le y_1\} = \{s \in S : Y(s) \le y_1\}$.

§2.2–2.4

5) * A RV is <u>discrete</u> if it can only assume a finite or countably infinite number of distinct values. The collection of these probabilities is the <u>probability distribution</u> of the discrete RV.

6) * For a discrete RV Y, the <u>probability mass function</u> (pmf) $p(y) = P(Y=y)$ for all y. Generally if $p(y) = 0$, that value of y is omitted.

7) For the function $p(y)$ to be a pmf for a discrete RV, the following must hold

i) $0 \le p(y) \le 1$ $\forall y$ (for all)

ii) $\sum_y p(y) = 1$ where the sum is over all values $y : p(y) > 0$.

8) <u>common E1 problem</u> give a table with a $p(y)$ omitted.

ex)

y	0	1	2	3
$p(y)$	.1	.2	.3	

Then $p(3) = .4$
$= 1 - .1 - .2 - .3$.

9) A probability distribution is a <u>model</u> or hopefully useful approximation for the data population.

that we want information about. objects

ex) pop = IQ scores of college graduates

" SJU "

10) Know p34 Let Y be a discrete RV with pmf $p(y)$. Then the expected value of Y is

$$E(Y) = \underbrace{\sum_y}_{y:\, p(y)>0} y \cdot p(y).$$

11) common problem Given a table of y and $p(y)$, find $E(Y)$.

ex)

y	0	1
$p(y)$	$\frac{1}{2}$	$\frac{1}{2}$

$E(Y) = \sum y\, p(y) = 0(\frac{1}{2}) + 1(\frac{1}{2}) = \boxed{\frac{1}{2}}$

ex)

y	-1	1
$p(y)$	$\frac{1}{2}$	$\frac{1}{2}$

$E(Y) = -1(\frac{1}{2}) + 1(\frac{1}{2}) = \boxed{0}$

ex)

y	-2	-1
$p(y)$	$\frac{1}{2}$	$\frac{1}{2}$

$E(Y) = -2(\frac{1}{2}) + -1(\frac{1}{2}) = \boxed{\frac{-3}{2}}$

12) $E(Y) = \underset{\text{mu}}{\mu}$ = population mean is a measure of location of the pop

ex) mean ht of adult US woman
wt man
networth
yearly income

and if $g(Y)$ is a real valued function of Y, then (14.5)

$g(Y)$ is a RV and $E\,g(Y) = E[g(Y)] = \underbrace{\sum_y}_{y:\,p(y)>0} g(y)\,p(y)$

14) know pHO The variance of a RV Y is

$V(Y) = \text{var}(Y) = E(Y - E(Y))^2$. The standard deviation

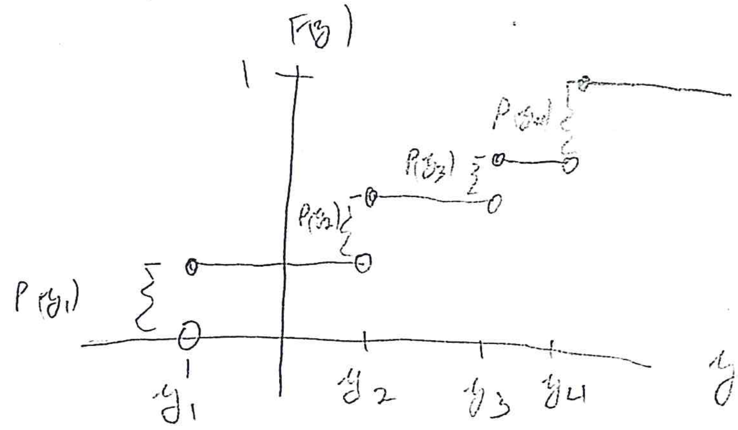
(SD) of Y is $SD(Y) = \sqrt{V(Y)}$. know that $V(Y) \geq 0$ and $SD(Y) \geq 0$.

15) $V(Y) = \sigma^2$, $SD(Y) = \sigma$ sigma

$V(Y) = \sum (y - E(Y))^2\, p(y)$ so $g(y) = (y - E(Y))^2$

16) common final problem! Given a table of y and $p(y)$, find $E[g(Y)]$, $V(Y)$ and $SD(Y)$.

17) If Y has pmf $p(y)$, then the cdf of Y is a step function with jumps of size $p(y)$ for values of y where $p(y) > 0$. See HW3 # 3.



$p(y) = \overbrace{F(y) - F(y-)}^{P(Y \leq y) - P(Y < y)} = P(Y = y)$

for $y: p(y) > 0$

y	y_1	y_2	y_3	y_4
$p(y)$	$p(y_1)$	$p(y_2)$	$p(y_3)$	$p(y_4)$

ex) $p(y)$ | 1 $\qquad p(y-E(Y))$ | 1

$E(Y) = 7(1) = 7 \qquad V(Y) = 0, \; SD(Y) = \sqrt{0} = 0$

ex) Quiz 3 may ask you to make a table to compute σ or σ^2.

y	$p(y)$	$y\,p(y)$	$y-E(Y)$	$(y-E(Y))^2 p(y)$
3	.3	.9	-1	0.3
4	.4	1.6	0	0
5	.3	1.5	1	0.3
$\Sigma p(y) = 1$		$\boxed{E(Y) = 4}$		$\boxed{0.6 = V(Y)} = \sum_y (y-EY)^2 p(y)$

$\Sigma (y - EY)$ is not useful but

$\Sigma (y-EY)\,p(y) = E(Y-EY) = 0$

$\sigma = SD(Y) = \sqrt{0.6} = 0.7746$

ex)

y	$p(y)$	$y\,p(y)$	$y-E(Y)$	$(y-E(Y))^2 p(y)$
1	.4	.4	-3	3.6
2	.1	.2	-2	0.4
6	.3	1.8	2	1.2
8	.2	1.6	4	3.2
$\Sigma p(y) = 1$		$\boxed{E(Y) = 4}$		$\boxed{8.4 = V(Y)} = \sum_y (y-EY)^2 p(y)$

$\sigma = SD(Y) = \sqrt{8.4} = 2.8983$

After making a table or two, you may want to say $\sigma^2 = V(Y) = \Sigma (y-EY)^2 p(y) =$

$(1-4)^2(.4) + (2-4)^2(.1) + (6-4)^2(.3) + (8-4)^2(.2) = 8.4$

Also short cut formula $V(Y) = E(Y^2) - (EY)^2$ is easier.

measure of spread (variability)

of the population

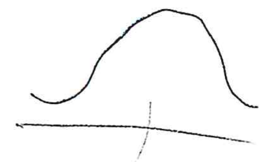
15.5

19) Empirical rule: For many bell shaped populations,

$\mu \pm \sigma$ contains $\simeq$ 68% of measurements

$\mu \pm 2\sigma$ 95%

$\mu \pm 3\sigma$ 99.7% $\simeq$ all



(histogram or pdf)

20)* Expectation rules:

$E(c) = c$ if $g(Y) \equiv c$ is a constant,

$E[c\,g(Y)] = c\,E[g(Y)]$

$E\left[\sum_{i=1}^{k} g_i(Y)\right] = \sum_{i=1}^{k} E[g_i(Y)]$

expected value of a sum is the sum of the expected values

21) p41 Know Short cut formula for variance

$$V(Y) = E(Y-EY)^2 = E(Y^2) - [E(Y)]^2.$$

Proof) Let $\mu = EY$. $V(Y) = E(Y-\mu)^2 = E[Y^2 - 2Y\mu + \mu^2]$

$= EY^2 - 2\mu E(Y) + \mu^2 = EY^2 - 2\mu^2 + \mu^2$

$= EY^2 - \mu^2.$

22) Know $E(Y^2) = V(Y) + [E(Y)]^2.$

y	y^2	$p(y)$	$y\,p(y)$	$y^2 p(y)$
1	1	.4	.4	.4
2	4	.1	.2	.4
6	36	.3	1.8	10.8
8	64	.2	1.6	12.8
			$E(Y)=4$	$24.4=E(Y^2)$

$V(Y)=E(Y^2)-[E(Y)]^2=24.4-16=8.4$ as before.

$SD(Y)=\sigma=\sqrt{8.4}$. Also

$$E(Y^2)=\sum y^2 p(y)=1^2(.4)+2^2(.1)+6^2(.3)+8^2(.2)$$
$$=24.4$$

23) *p26 An expt is a binomial experiment if

1) the expt consists on n identical trials
2) each trial has 2 outcomes = "success" and F
3) For each trial, P(Success) $=p$, $P(F)=1-p$.
4) The trials are independent.
5) RV Y = number of successes in n trials

Note: "sucess" does not have the dictionary meaning, it is what you count, eg males, females number of people that die within 5 years of being diagnosed with colon cancer.

ex) Flip coin n times; count # of heads eg polls with random samples of size n

Y counts # for Clinton. This expt is approx binomial (the trials are approx ind if the sample size $n << N$ = pop size).

Similarly randomly select 1000 people, Y counts # that test HIV positive

ex) A voluntary response sample consists of people who choose themselves by responding to a general appeal. People with strong opinions are likely to respond and the sample does not represent a pop and not ind, so not a binom expt. eg Ann Landers asked her readers if they had it to do over again would they have kids (after getting a letter from a parent with horrible kids). Almost 10000 parents responded and 70% said no.

verbally

american idol

the voice

nonscientific poll

In a random sample ≈ binom expt, 9% would say no.

ex) prof needs student to comment on his teaching and picks several A students. Y = # that give high rating. This is not a bin expt (pop of prof students is not represented, not ind)

16.5

4) know RV Y has a binomial distribution based on n trials with success prob p, written $Y \sim bin(n,p)$, if the pmf

$\binom{n}{y} p^y (1-p)^{n-y}$ for $y = 0, 1, \ldots, n$. The pmf formula is for $p \in (0,1)$.

p27
Note: $q = 1-p$. If there are n trials and y S's then there are $n-y$ F's.

P(y S's and n−y F's) $= p^y (1-p)^{n-y}$ by ind.

There are $\binom{n}{y}$ sequences with y S's. By the binomial th, $1 = (p + 1 - p)^n = \sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y}$

So $p(y)$ is a pmf.

25) Know! If $Y \sim bin(n,p)$, $E(Y) = np$ and $V(Y) = np(1-p)$.

common error! try to find $E(Y) = \sum y\, p(y)$ and $V(Y) = \sum (y - EY)^2 p(y)$ when $Y \sim bin(n,p)$. The above formulas are much easier.

p26
26) Know A RV X is a Bernoulli RV if $X \sim bin(n=1, p)$

Note: p 35 proof that $E(Y) = np$ for $Y \sim bin(n,p)$

$$E(Y) = \sum y\, p(y) = \sum_{y=0}^{n} y \frac{n!}{y!\,(n-y)!} p^y (1-p)^{n-y} =$$

$$np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!\,(n-y)!} p^{y-1} (1-p)^{n-y}$$

$y=0$ term is 0
$\frac{y}{y!} = \frac{1}{(y-1)!}$
$n! = n(n-1)!$

Let $z = y - 1$ then $y = n \to z = n-1$, $y = 1 \to z = 0$ } like U substitution

$$\text{and } E(Y) \overset{\text{trick}}{=} np \sum_{z=0}^{n-1} \frac{(n-1)!}{z!\,(n-1-z)!} p^z (1-p)^{n-1-z} =$$

$\sum_{z=0} \ldots = np$ ← trick

$1 = \sum_z P_{bin(n-1,p)}(z)$

175

ex3 Suppose $Y \sim bin(n,p)$ where Y counts the number of times where outcome D occured where $P(D) = p$ for a single experiment.

better for ex and

Find i) P(D occurred in <u>none</u> of the n expts)

ii) P(D occurred in <u>at least one</u> of the n expts)

iii) P(D occurred in <u>all</u> n expts)

iv) P(D occurred in <u>not all</u> n expts)

✓ ↔ D occurs, blank ↔ D did not occur

at least one

none

P(at least one) = 1 − P(none)

soln) i) $P(\text{none}) = P(Y=0) = \binom{n}{0} p^0 (1-p)^n = (1-p)^n$

complement rule

$\underset{1}{\underline{1-p}} \quad \underset{2}{\underline{1-p}} \quad \cdots \quad \underset{n}{\underline{1-p}}$ trial

$Y \in \{0,1,\ldots,n\}$

ii) $P(\text{at least one}) = P(Y \geq 1) = 1 - P(Y<1) = 1 - P(Y=0)$

$= P(\text{not none}) = 1 - P(\text{none}) = 1-(1-p)^n = 1 - \binom{n}{0} p^0 (1-p)^n$

iii) $P(\text{all}) = P(Y=n) = \binom{n}{n} p^n (1-p)^0 = p^n$

$\underset{1}{\underline{p}} \quad \underset{2}{\underline{p}} \quad \cdots \quad \underset{n}{\underline{p}} = p^n$ trial

iv) $P(\text{not all}) = 1 - P(\text{all}) = 1 - P(Y=n) = 1 - \binom{n}{n} p^n (1-p)^0 = 1-p^n$

27) $P(Y \text{ is at most } k) = P(Y \leq k)$, $P(Y \text{ is at least } k) = P(Y \geq k)$

$P(Y \text{ is at most } k) = P(Y \le k) = p(0) + \cdots + p(k)$

$= 1 - P(Y > k) = 1 - P(Y \ge k+1) = 1 - [p(k+1) + p(k+2) + \cdots + p(n)]$.

$P(Y \text{ is at least } k) = P(Y \ge k) = p(k) + \cdots + p(n)$

$= 1 - P(Y < k) = 1 - P(Y \le k-1) = 1 - p(0) - \cdots - p(k-1)$,

ex) common E1 problem: Fair die is rolled 3 times. Y counts # of 5's. Then $Y \sim \text{bin}(n=3, p=\frac{1}{6})$

y	0	1	2	3
$p(y)$	.5787	.3472	.0694	.0046

(sums to 0.9999 due to rounding)

a) $E(Y) = np = 3\,\frac{1}{6} = \boxed{0.5}$

b) $V(Y) = np(1-p) = 3\,\underbrace{\frac{1}{6}\,\frac{5}{6}}_{\text{sum to } 1} = \frac{5}{12} = \boxed{0.4167}$

c) $P(\text{none of the rolls are 5's}) = P(Y=0) = p(0) = \binom{3}{0}\left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3$ (sum to)

$= \left(\frac{5}{6}\right)^3$

d) $P(\text{at least one roll is a } 5) = P(Y \ge 1) = p(1) + p(2) + p(3)$

$= 1 - p(0) = 1 - \left(\frac{5}{6}\right)^3$.

e) $P(\text{all 3 rolls are 5's}) = P(Y=3) = p(3) = \binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0$

$= \left(\frac{1}{6}\right)^3$

f) $P(\text{not all 3 rolls are 5's}) = P(Y \ne 3) = 1 - P(Y=3)$

$= 1 - p(3) = p(0) + p(1) + p(2) = 1 - \left(\frac{1}{6}\right)^3$.

$= \binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 + \binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \left(\frac{5}{6}\right)^3 + 3\frac{1}{6}\left(\frac{5}{6}\right)^2 =$

$0.5787 + 0.3472 = \boxed{0.9259}$

h) P(at least 2 rolls are 5's) $= P(Y \geq 2)$

$= P(2) + P(3) = 1 - P(0) - P(1) = 1 - .9259 = \boxed{0.0741}$

28) *common error: $q = 1-p$ is given in the story problem and student uses $Y \sim bin(n, q)$ instead of $Y \sim bin(n, p)$, see HW 3 #6.

29) * p29-30 A poisson(λ) RV Y is used to model the total # of occurrences of some phenomena in a fixed period of time or a fixed region of space where λ is the average value of Y.

eg # telephone calls in a 5 minute interval (at a center with several phones)

\# radioactive particles that hit a target in 3 hours

30) p29 know A RV Y has a poisson(λ) dist if Y has pmf $p(y) = \dfrac{\lambda^y e^{-\lambda}}{y!}$, $y = 0, 1, 2, \ldots$

for $\lambda > 0$.

31) know p36 If $Y \sim$ Poisson(λ) then $E(Y) = V(Y) = \lambda$.
(is distributed as)