MARCH 12 /18
APPLIED ANAL.

Last time
- Monty-hall
- Discrete random variable
- discrete probability distribution
- $f(x) \geq 0$, $\sum_{all\,x} f(x) = 1$
- Cumulative distribution $F(x)$
- Bernouli trials
- Binomial distribution $\binom{n}{x} p^x (1-p)^{n-x}$

- e.g. you take a test with 13 derivative questions, 7 integral questions. Each question has 5 possible answers. You feel that you have a 90% chance on each derivative question, you must guess on the integral questions. You're offered \$500 if you score 95% or better. Find your probs. of winning.

$Pr\left(\frac{13}{13}\text{ deriv.}, \frac{6}{7}\text{ int}\right) + Pr\left(\frac{12}{13}\text{ deriv.}, \frac{7}{7}\text{ integral}\right) + Pr\left(\frac{13}{13}\text{ deriv.}, \frac{7}{7}\text{ int}\right)$

$= (.9)^{13}\binom{7}{6}(.2)^6(.8)^1 + \binom{13}{12}(.9)^{12}(.1)^1(.2)^7 + (.9)^{13}(.2)^7$

(*) The <u>cumulative binomial distribution</u> is

$B(x,n,p) = \sum_{i=0}^{x} b(i,n,p)$

There is a table - You may not get it in the exam

- e.g. You perform 10 Bernouli trials with $p = .4$. In terms of B, Find:

(i) The prob of getting at most 3 successes
(ii) " " fewer than 3 successes
(iii) " " more than 3 successes
(iv) " " at least 3 successes
(v) " " at least 3, no more than 5 successes
(vi) " " 3 successes

(i) $B(3, 10, .4)$ (ii) $B(2, 10, .4)$ (iii) $-B(3, 10, .4)$
(iv) $1-B(2, 10, .4)$ (v) $B(5, 10, .4) - B(2, 10, .4)$
(vi) $B(3, 10, .4) - B(2, 10, .4)$

<u>Hypergeometric distribution</u>: We have N items and a are special. We randomly select n items without replacement. The prob. of getting x special items in our sample is:

$$L(x; n, a, N) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

- e.g. Find the prob of getting exactly two clubs in a particular hand

$$\frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}}$$

- e.g. Suppose we have 1000 batteries 70 of which are dead. If we randomly select 40 batteries Find the prob. of getting 8 dead ones if:
  (i) We remove each battery after test
  (ii) We replace each dead battery

→ (i) $\frac{\binom{70}{8}\binom{930}{32}}{\binom{1000}{40}}$ (ii) $\binom{40}{8}\left(\frac{70}{1000}\right)^8\left(\frac{930}{1000}\right)^{32}$

As $n \ll N$, the hypergeometric can be approximated by a second l - general, the approximation is good at $n \leq \frac{N}{(a}$

- e.g. We have a large collection of light bulbs, 30% of which are burnt out. We randomly select 100 bulbs, find the Prob. of getting 25 burnt out bulbs. We don't know N, must approx. using binomial

$$\binom{100}{25}(.3)^{25}(.7)^{75}$$

(Sections 2.5, 2.6)

Suppose we have numbers $x_1, x_2, \ldots, x_n$

The <u>Sample mean</u> is $\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$

- e.g. 1, 7, 3, 2 $\bar{x} = \frac{1+7+3+2}{4} = \frac{13}{4}$

If $x_1 \le x_2 \le x_3 \le \dots \le x_n$, then the sample median

$$\begin{cases} x_{(n+1)/2}, & n \text{ odd} \\ \dfrac{x_{n/2} + x_{(n+1)}}{2}, & n \text{ even} \end{cases}$$

- e.g. 2, 5, 3, 1, 8, 14, 17

  1 2 3 5 8 14 17 $\rightarrow$ 5 median

- e.g. 2 3 8 1 4 6

  1 2 3 4 6 8 $\rightarrow$ $\frac{3+4}{2} = 7/2$ median

- e.g. 60 60 60 60 60 60 } 60 median

  20 20 20 100 100 100

The <u>sample variance</u> is $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2}{n-1}$

- e.g. 1, 3, 4, 8 $\bar{x} = \dfrac{1+3+4+8}{4} = 4$

$$s^2 = \frac{(1-4)^2 + (3-4)^2 + (4-4)^2 + (8-4)^2}{3} = \frac{26}{3}$$

This measures how spread out the values are

The <u>sample standard deviation</u> is $s = \sqrt{s^2}$

- e.g. $s = \sqrt{26/3}$

Suppose we have $x_1 \le x_2 \le x_3 \dots \le x_n$

Let $0 < p < 1$. Then the $(100p)^{th}$ <u>percentile</u> is a number q so that

(a) at least $100p\%$ of the data is $\le q$ and

(b) at least $100(1-p)\%$ of the data is $\ge q$

To find the $(100p)^{-th}$ percentile, calculate np

If np is not an integer, round up to the next <u>larger</u> integer, use x_k

If np is an integer, use $\dfrac{x_{np} + x_{np+1}}{2}$

- e.g. 1, 2, 8, 11, 14, 19, 23, 27, 30, 42

Find (i) 23rd percentile (ii) 80th percentile

(i) $np = 10(.23) = 2.3$, $x_3 = 8$

(ii) $np = 10(.8) = 8$, $\frac{x_8 + x_9}{2} = \frac{27+30}{2} = \frac{57}{2}$

The quartiles are the 25th, 50th, 75th percentiles

Q_1 : 25th

Q_2 : 50th - median

Q_3 : 75th

- e.g. 1, 8, 2, 7, 3, 5

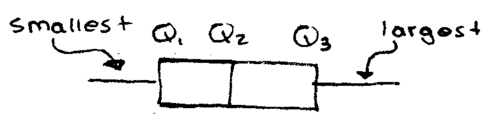
1, 2, 3, 5, 7, 8

Q_1 : $np = 6(.25) = 1.5$, $x_2 = 2$

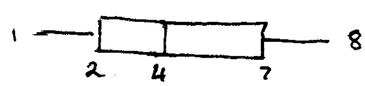
Q_2 : $\frac{3+5}{2} = 4$ (median)

Q_3 : $np = 6(.75) = 4.5$, $x_5 = 7$

A <u>boxplot</u> can be used to represent the data



$\Longrightarrow$



The <u>interquartile range</u> is $Q_3 - Q_1$

- e.g. $7 - 2 = 5$

MARCH 14TH/18
Applied Analysis

- Last time - cumulative distribution
  - hypergeometric $\dfrac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$
  - Sample mean, median
  - sample variance, sample standard deviation
  - percentiles, quartiles
  - box plot, interquartile range

Let x be a discrete random variable, then is <u>mean</u> (or <u>expected value</u>) is $\mu = E(x) = \sum_{\text{all } x} x f(x)$

- e.g.

x	1	2	3
$f(x)$	$1/2$	$1/4$	$1/4$

$\mu = 1(1/2) + 2(1/4) + 3(1/4) = 7/4$

We say that a gain is Fair if our expected gain is $\emptyset$

- e.g. In a lottery 1000 tickets are sold. 1 First prize of \$500, and there are 10 prizes of \$100. How much should a ticket cost to make it fair?

Let c be a cost of a ticket, Let x be our

x	$500-c$	$100-c$	$-c$
$f(x)$	$\frac{1}{1000}$	$\frac{10}{1000}$	$\frac{989}{1000}$

$\emptyset = \mu = (500-c)\left(\frac{1}{1000}\right) + (100-c)\left(\frac{10}{1000}\right) - c\left(\frac{989}{1000}\right)$

$\emptyset = 1500 - 1000c, \quad c = 50$

- e.g. In chuck-a-luck, you pick a number between 1 → 6. Three balanced dice are rolled. If your number comes up at least once, you gain a dollar for each time it appears. If your number doesn't come up, you lose a dollar. Find expected gain/loss.

x	3	2	1	-1
$f(x)$	$(1/6)^3$	$\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$ $= \frac{15}{6^3}$	$\binom{3}{1}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2$ $= 75/6^3$	$(5/6)^3$ $= \frac{125}{216}$

$\mu = 3\left(\frac{1}{216}\right) + 2\left(\frac{15}{216}\right) + 1\left(\frac{75}{216}\right) - 1\left(\frac{125}{216}\right) = \frac{-17}{216}$

	1	2	3	4	5	6	
123 :	-1	-1	-1	+1	+1	+1	0
112 :	-2	-1	+1	+1	+1	+1	1
111 :	-3	+1	+1	+1	+1	+1	2

For a binomial : $\mu = np$

- e.g. Flip a balanced coin 80 times, expected # of heads : $80(1/2) = 40$

For hypergeometric : $\mu = n\left(\frac{a}{N}\right)$

- e.g. We have 100 bottles of water, 20 of which are poisoned. If we randomly select 8 bottles, the expected number of poisoned ones is $8\left(\frac{20}{1000}\right) = 1.6$

We can measure how spread out the values of x tend to be

The variance is $\sigma^2 = \sum_{\text{all } x} (x-\mu)^2 f(x)$

- e.g.

x	4	8	12
$f(x)$	$1/4$	$1/2$	$1/4$

$\mu = 4(1/4) + 8(1/2) + 12(1/4) = 8$

$\sigma^2 = (4-8)^2(1/4) + (8-8)^2(1/2) + (12-8)^2(1/4)$

$= 8$

The standard deviation is $\sigma = \sqrt{\sigma^2}$

- e.g. above, $\sigma = \sqrt{8}$

For binomial; $\sigma^2 = npq = np(1-p)$

For hypergeometric; $\sigma^2 = n\left(\frac{a}{N}\right)\left(-\frac{a}{N}\right)\left(\frac{N-n}{N-1}\right)$

- e.g. We have 1000 cars, 200 are green. If we randomly select 80 cars, Find the standard dev. in the number of cars if : (i) we sample w/o replacement

(ii) we replace each after sampling

(i) $\sqrt{80\left(\frac{200}{1000}\right)\left(\frac{800}{1000}\right)\left(\frac{920}{999}\right)}$ (ii) $\sqrt{80(.2)(.8)}$

<u>Chebyshev's Theorem:</u> Let x be a random variable with mew μ, standard deviation σ. Then, for any $k > 0$,

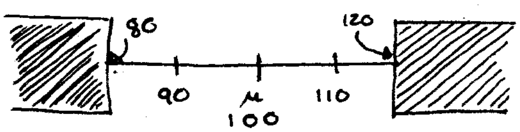
$$\Pr\left[|x-\mu| \geq k\sigma\right] \leq 1/k^2$$

Proof:

$$\sigma^2 = \sum_{\text{all } k} (x-\mu)^2 f(x)$$
$$\geq \sum_{|x-\mu| \geq \sigma k} (x-\mu)^2 f(x)$$
$$\geq \sum_{|x-\mu| \geq \sigma k} k^2 \sigma^2 f(x)$$

$$\therefore\ 1 \geq \sum_{|x-\mu| \geq k\sigma} k^2 f(x)$$
$$1/k^2 \geq \sum_{|x-\mu| \geq k\sigma} f(x)$$
$$\Pr\left(|x-\mu| \geq k\sigma\right) \leq 1/k^2$$

- e.g. X has mean 100 and standard deviation 10
What can we say about the prob that $x \leq 80$ or $x \geq 120$



$\mu = 100$

$\mu + k\sigma = 120$

$\sigma = 10$; $k = 2$

$$\Pr\left[(x-100) \geq 2(10)\right) \leq 1/2^2 = 1/4$$

Prob is at most $1/4$

$$\Pr(80 \leq x \leq 120) = 1 - \Pr(x \leq 80 \text{ or } x \geq 120) \geq 1 - 1/4 = 3/4$$

MARCH 16/18
APPLIED ANAL.

- last time: - $\mu = \sqrt{2}\,|x|$,
  - binomial, hypergeometric
  - variance σ^2, standard deviation σ
  - Chebyshev's thm: $Pr(|x-\mu| \geq k\sigma) \leq \frac{1}{k^2}$
- e.g. we flip a balanced coin 10000 times, find out what Chebyshev says about the prob. of the proportion of heads being at most 45% or at least 55%

X = # of heads, $n = 10000$, $p = 1/2$

$\left|\frac{X}{n} - .5\right| \geq .05$

$|x - \underbrace{.5n}_{\mu}| \geq \underbrace{.05n}_{k\sigma}$

x is binomial, $\mu = np = .5n = 5000$

$k\sigma = .05n$, but $\sigma = \sqrt{np(1-p)} = \sqrt{n/4}$

$k\sigma = k(\sqrt{n/4}) = .05n$

$k = \frac{.05n}{\sqrt{n/4}} = \frac{.05\sqrt{n}}{1/2} = .1\sqrt{n} = .1(100) = 10$

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

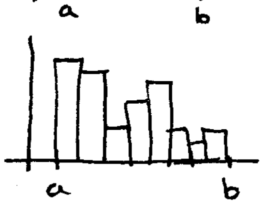
prob $\leq 1/10^2 = .01$

Chapter 5 - Probability Densities (5.1 - 5.6)

A <u>continuous random variable</u> X has values in $\mathbb{R}$ or some interval. We will not worry about the probability of x being a particular value. Instead, we look at the prob of x being in some interval.

Let us say that x takes on values in $[a, b]$

Area of each rectangle = prob. that x is in that interval



Refine
repeat

In the limits we assign a function $f(x)$ so that
$P_r[c \le x \le d] = \int_c^d f(x)dx$

We call $f(x)$ the probability density function

Rules: (i) $f(x)$ is an $\int$-able function (?)

(ii) $f(x) \ge 0$ for all x

(iii) $\int_{-\infty}^{\infty} f(x)dx = 1$

If $f(x)$ is only defined on $[a,b]$, make it 0 everywhere else

- e.g. X has density function $f(x) = \begin{cases} Kx^3, & x \in [0,1] \\ 0, & \text{elsewhere} \end{cases}$

(i) Find K (ii) Find $P_r(1/4 \le x \le 1/2)$ (iii) Find $P_r(x < 1/3)$

(iv) Find $P_r(x > 2/3)$ (v) Find $P_r(x = 1/2)$

→ (i) $1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^1 Kx^3 dx = \frac{Kx^4}{4}\Big|_0^1 = \frac{K}{4}$, so $K = 4$

(ii) $\int_{1/4}^{1/2} f(x)dx = \int_{1/4}^{1/2} 4x^3 dx = x^4\Big|_{1/4}^{1/2} = 1/16 - 1/256$

(iii) $\int_{-\infty}^{1/3} f(x)dx = \int_0^{1/3} 4x^3 dx = x^4\Big|_0^{1/3} = 1/81$

(iv) $\int_{2/3}^{\infty} f(x)dx = \int_{2/3}^1 4x^3 dx = x^4\Big|_{2/3}^1 = 1 - 16/81$

(v) 0

- e.g. x has density function $f(x) = \begin{cases} 3e^{-3x}, & x > 0 \\ 0, & x \le 0 \end{cases}$

(i) Find $P_r(x < 4)$ (ii) Find $P_r(x > 2)$

→ (i) $\int_{-\infty}^4 f(x)dx \Rightarrow \int_0^4 3e^{-3x} \Rightarrow -e^{-3x}\Big|_0^4 = -e^{-12} + 1$

(ii) $\int_2^{\infty} f(x)dx \Rightarrow \int_2^{\infty} 3e^{-3x} \Rightarrow -e^{-3x}\Big|_2^{\infty} = 0 - (-e^{-6}) = e^{-6}$

The mew of x is $\mu = \int_{-\infty}^{\infty} x f(x)dx$

The variance is $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

The standard deviation is $\sigma = \sqrt{\sigma^2}$

- e.g. $f(x) = \begin{cases} 4x^3, & x \in [0,1] \\ 0, & \text{elsewhere} \end{cases}$

$\mu = \int_{-\infty}^{\infty} x f(x)dx = \int_0^1 4x^3 dx = \frac{4x^5}{5}\Big|_0^1 = 4/5$

$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_0^1 (x - 4/5)^2 (4x^3)dx$

$\ldots \Rightarrow 4/6 - \frac{32}{25} + \frac{64}{100}$

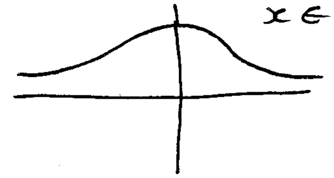
The distribution factor is $F(x) = Pr(X \le x) = \int_{-\infty}^{x} f(t)\,dt$

- e.g. In the above example,
  $F(x) : \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} 4t^3\,dt = t^4 \Big|_0^x = x^4, \; x \in [0,1]$
  IF $x < 0$, $F(x) = 0$. IF $x > 1$, $F(x) = 1$

x has normal distribution with mew μ, standard deviation σ if it's density Function is:

$x \in \mathbb{R}$

$$f(x) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) e^{-(x-\mu)^2/(2\sigma^2)}$$



This Function cant be integrated, but we can approximate as closely as needed.

Z is standard normal if it is normal with $\mu = 0$, $\sigma = 1$

We have the distribution Function $F(z) = Pr(Z \le z)$ in a table.

- e.g. Find (i) $Pr(Z < 1.36)$ (ii) $Pr(Z > -0.82)$
  (iii) $Pr(0.34 \le Z \le 1.20)$

→ (i) $F(1.36) = 0.9131$

(ii) $1 - Pr(Z < -0.82) \Rightarrow 1 - F(-0.82)$
$\Rightarrow 1 - 0.2061 = 0.7939$

(iii) $Pr(Z < 1.20) - Pr(Z < 0.34)$
$\Rightarrow F(1.20) - [Pr(Z < 0.34)]$
$\Rightarrow F(1.20) - [F(0.34)]$
$\Rightarrow 0.8849 - 0.6331 = 0.2518$

- e.g. Find (i) $Pr(-1 \le Z \le 1)$ (ii) $Pr(-2 \le Z \le 2)$
  (iii) $Pr(-3 \le Z \le 3)$

(i) $Pr(Z < 1) - Pr(Z < -1) = F(1) - F(-1) = 0.6826$
(ii) $Pr(Z < 2) - Pr(Z < -2) = F(2) - F(-2) = 0.9544$
(iii) $Pr(Z < 3) - Pr(Z < -3) = F(3) - F(-3) = 0.9974$

- e.g. Find a so that $Pr(Z > a) = .26$
  $Pr(Z \le a) = .74 \Rightarrow F(.64) = .7389, \; F(.65) = .7454$
  $a = .645$