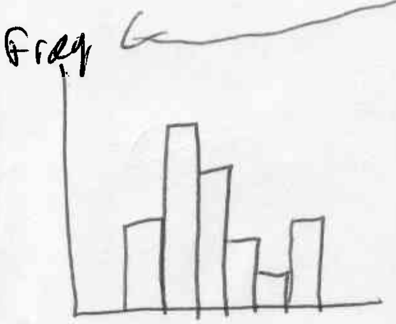
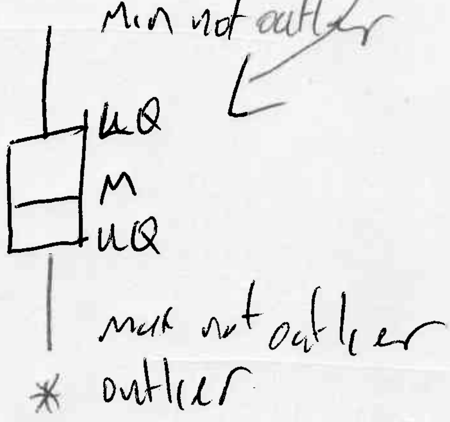
Math 324 Review for midterm

What have we covered?

- Descriptive Statistics
  - graphical methods
    - stem and leaf
    - histogram
    - boxplot
  - Summary Statistics
    - measures of location
      - mean $\bar{x} = \frac{\sum x_i}{n}$
      - median
      - Quartiles (LQ, UQ)
      - trimmed mean
    - measures of variability
      - variance $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
      - std deviation $s = \sqrt{s^2}$
      - $IQR = UQ - LQ$

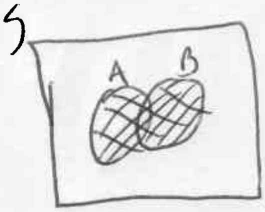
```
1 |
2 | 9
3 | 0
4 | 3 3 7 9
5 | 1 2 2 4 5 5 8 8 9
6 | 1 1 3
7 |
```

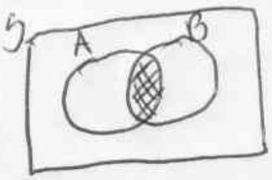


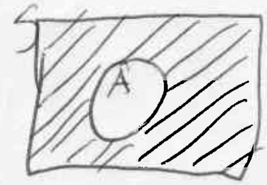


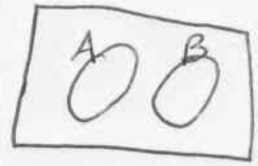
- Probability

- sample space S
- events
- set theory









  - $A \cup B$
  - $A \cap B$
  - A^C
- mutually exclusive/disjoint

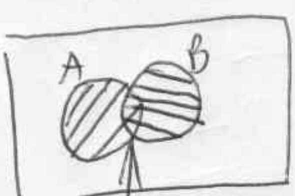
- Probability Axioms
  - $P(A) \geq 0$ for any event A
  - $P(S) = 1$
  - $P(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} P(A_i)$ if all A_i M.E.

- other probability properties

$P(A) = 1 - P(A^C)$

$P(A \cap B) = 0$ if A, B M.E.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$



counting this twice in $P(A)$, $P(B)$ so subtract one.

- Equally likely outcomes

$$p = \frac{1}{N} \text{ for any simple event}$$

$$P(A) = \frac{N(A)}{N}$$

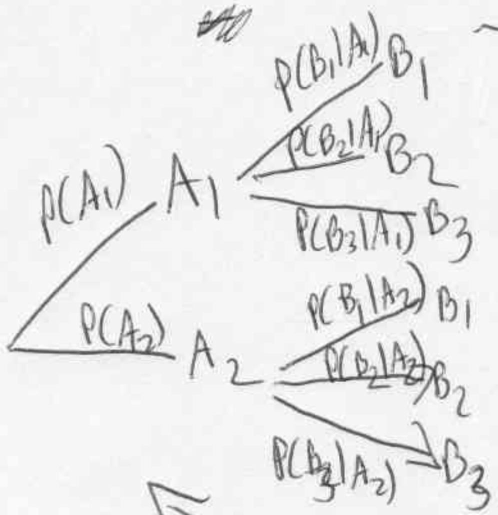
- Counting rules
  - ordered pairs $n_1 n_2$
  - k tuples $n_1 n_2 \cdots n_k$
  - permutations (order matters) $^nP_k = \frac{n!}{(n-k)!}$
  - combinations (order does not matter) $^nC_k = \frac{n!}{(n-k)!\,k!}$

- Conditional probability

"probability of A given B happened" $P(A|B) = \frac{P(A \cap B)}{P(B)}$

- multiplication rule

$$P(A \cap B) = P(A|B)P(B)$$



- Bayes theorem
  - law of total probability

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

  - $$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

- tree diagrams

- Independence

A and B independent $\iff P(A \cap B) = P(A)P(B)$

$P(A|B) = P(A) \Rightarrow B, A$ independent

- Discrete Random variables
  - Random variable - numerical ~~value~~ based on outcome of random event

upper case eg X, Y

  - Discrete r.v.s - finite or countably infinite number of possible values

  - probability distribution

$$p(x) = P(X=x) = P(\forall s \in S : X(s) = x)$$

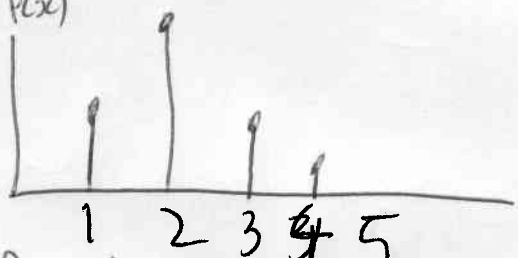
with $p(x) \geq 0$ $\quad \sum p(x) = 1$

5

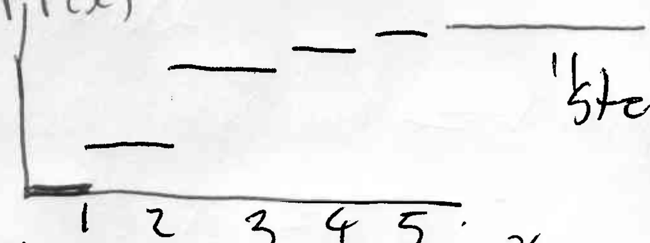
- cdf (cumulative distribution)

$$F(x) = P(X \le x) = \sum_{y: y \le x} p(y)$$

- graphically
  - pdf $p(x)$



  - cdf $F(x)$



"step function"

- relationship between pdf and cdf

$$P(a \le X \le b) = F(b) - F(a^-)$$

$$P(X = a) = F(a) - F(a-1)$$

- Expected Value

$$\mu_x = E[X] = \sum x_i p(x_i) \qquad E[h(X)] = \sum h(x_i) p(x_i)$$

$$Var(X) = \sum (x_i - \mu_x)^2 p(x_i)$$

or $Var(x) = E[x^2] - (E(x))^2$

quicker

- Binomial R.V
    - Fixed n trials
    - S or F each trial
    - independent
    - $P(S) = p$ constant

  sampling with replacement situations

    - $X \equiv$ "number of S in n trials"
    - pdf, mean, variance

$$b(x; n, p) = \begin{cases} \binom{n}{x} p^{x} (1-p)^{n-x} & x = 0, 1, \ldots n \\ 0 & \text{o.w} \end{cases}$$

$$E[X] = np$$

$$var(X) = np(1-p)$$

- Hypergeometric R.V
    - finite population N
    - M successes in population
    - sample $n < N$ without replacement

  sampling without replacement situations

$X =$ "number of S in sample"

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad \max(0, n-M+N) \le x \le \min(n, M)$$

$$E(X) = n\frac{M}{N} \quad Var(X) = \left(\frac{N-n}{N-1}\right) n\frac{M}{N}\left(1-\frac{M}{N}\right)$$

- can approximate Hypergeometric with binomial under certain conditions.

- Negative binomial
  - independent trials
  - S or F
  - $P(S) = p$ constant
  - stop when acheive r successes

X = "number of failures before rth success"

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

$$E[X] = \frac{r(1-p)}{p} \quad Var(X) = \frac{r(1-p)}{p^2}$$

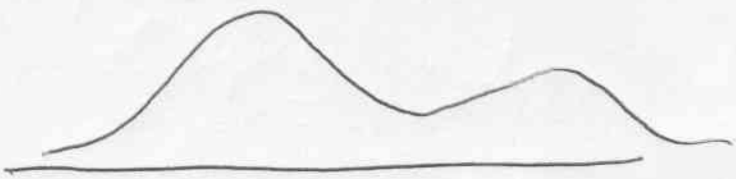
- Poisson

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \ldots$$

$$E[X] = Var(X) = \lambda$$

- poisson can be used to approx binomial in certain conditions.

- Continuous rv
  - some interval $A < B$ such that every value in $[A, B]$ is possible value of X.
  - pdf $f(x)$ density curve

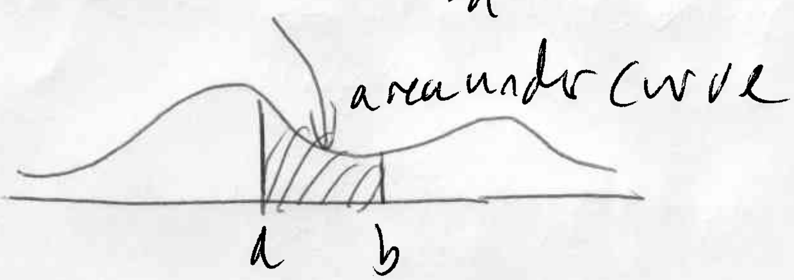


$f(x) \geq 0 \ \forall x$

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

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

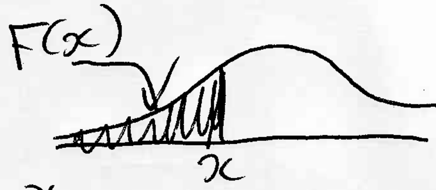
Note $P(X=a) = 0$ no area under curve above a single point.



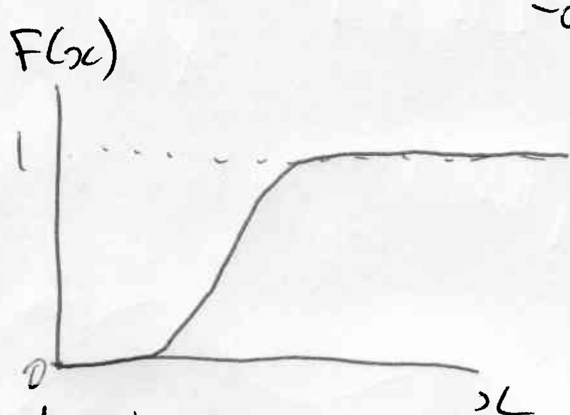
- uniform Distribution

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{o.w} \end{cases}$$

– cdf



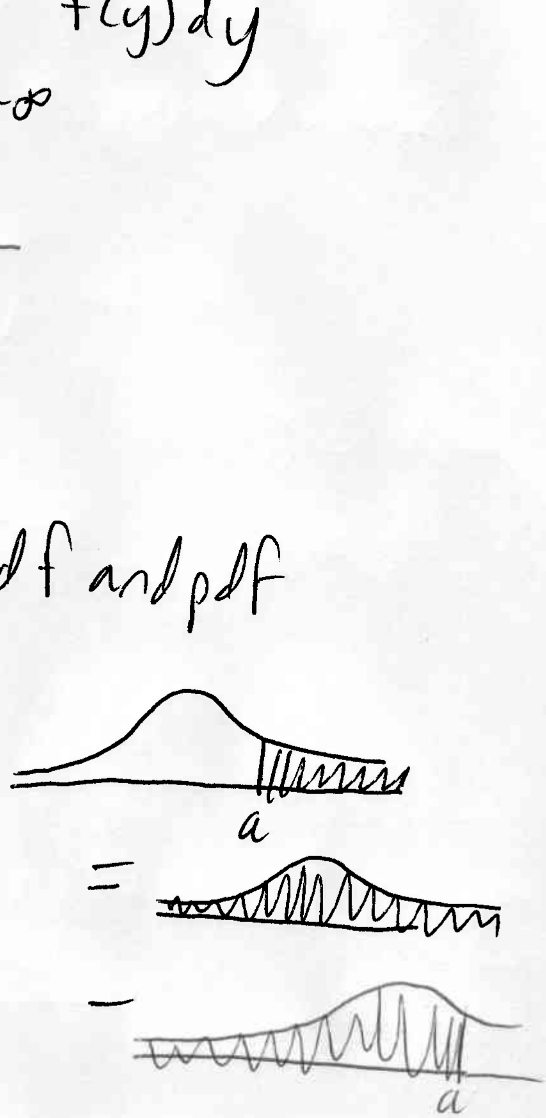
$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\,dy$$



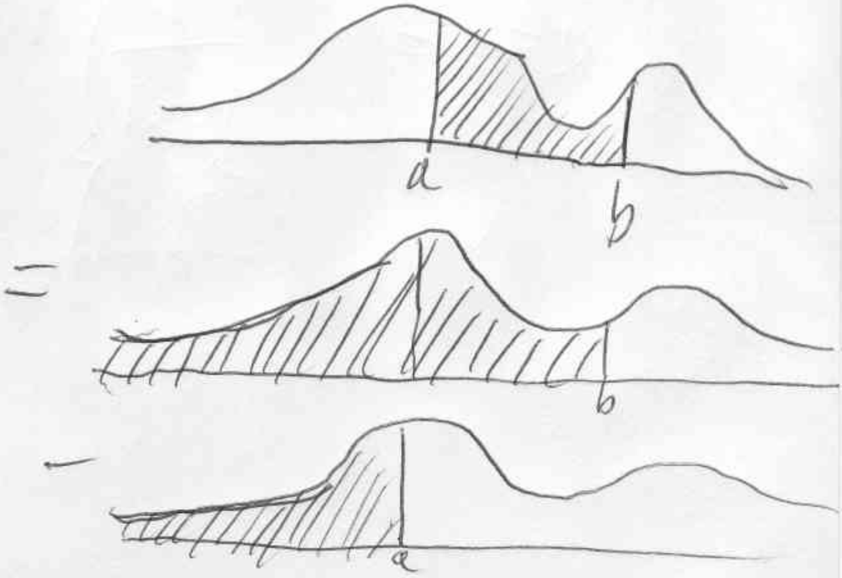
– relationships between cdf and pdf

$$\frac{d}{dx} F(x) = f(x)$$

$$P(X > a) = 1 - F(a)$$



$$P(a \le X \le b) = F(b) - F(a)$$



- percentile

  solution x_p of

  $$p = F(x_p)$$

- Expect value

  $$\mu_X = E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

  $$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\,dx$$

- Variance

  $$Var(X) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f(x)\,dx$$

  or $\underbrace{Var(X) = E[X^2] - (E(X))^2}_{\text{usually easier this way}}$

- Normal distribution
  - areas under curve using table.
  - percentiles using table given areas
  - standardization $z = \frac{X-\mu}{\sigma}$
  - unstandardization $\sigma z + \mu = X$

- normal can be used to approx binomial
  - $\mu = E[X] = np \quad \sigma = SD(X) = \sqrt{np(1-p)}$
  - continuity correction

- Gamma Distribution
  - special cases
    - exponential
    - χ^2 Chi-squared dist.

What would it be wise to review before midterm?

- Homework problems – look at solutions try to follow what is going on. If you made a mistake try to figure out why
- Worked examples in lectures – review, make sure that you understand the working in each problem (ie why is something done)

- Specific items you should make sure you understand well.
  - Normal distribution
    - areas under curve
    - percentiles
    - non std distributions.
  - Probability of a union eg $P(A \cup B) = ?$
  - independence
  - Hypergeometric distribution
  - poisson, binomial distribution
  - pdf, cdf, $E[X]$, $Var[X]$, percentile for r.v.