Tentative midterm date:
Thursday, October 18

Sorry for confusion about early November
Summary: Parameters of a Probability Distribution

- Probability Mass Function (PMF): Prob(X=x)
- Cumulative Distribution Function (CDF): Prob(X≤x)
- Complementary Cumulative Distribution Function (CCDF): Prob(X>x)
- The arithmetic mean or average value, $\mu=E[X]$, is a measure of the center of mass of a random variable.
- The variance, $V(X)=E[(X-\mu)^2]$, is a measure of the dispersion of a random variable around its mean. It has units of $\mu^2$.
- The standard deviation, $\sigma=V(X)^{0.5}$ is another measure of the dispersion around mean. It has the same units as $\mu$.
- The skewness, $\gamma_1=E[(X-\mu)^3/\sigma^3]$, measures the asymmetry around mean. When $>0$ fatter tails above mean. When $<0$ – fatter tails below mean.
- The geometric mean, $\exp(\mathbb{E}[\log X])$, is useful for very broad distributions.
A gallery of useful discrete probability distributions
Discrete Uniform Distribution

• Simplest discrete distribution.
• The random variable $X$ assumes only a finite number of values, each with equal probability.
• A random variable $X$ has a discrete uniform distribution if each of the $n$ values in its range, say $x_1, x_2, ..., x_n$, has equal probability.

$$f(x_i) = 1/n$$
Uniform Distribution of Consecutive Integers

• Let $X$ be a discrete uniform random variable all integers from $a$ to $b$ (inclusive). There are $b - (a-1)$ integers. Therefore each one gets:

$$f(x) = 1/(b-a+1)$$

• Its measures are:

$$\mu = E(x) = (b+a)/2$$
$$\sigma^2 = V(x) = [(b-a+1)^2-1]/12$$

Note that the mean is the midpoint of $a$ & $b$. 
Matlab exercise: Uniform distribution

• Generate a sample of size 1000 for uniform random variable X taking values 1,2,3,...,10
• Plot the approximation to the probability mass function based on this sample
• Calculate the mean and variance of this sample and compare it to theoretical calculations:
  \[ E[X] = \frac{a+b}{2} \] and \[ V[X] = \frac{(b-a+1)^2-1}{12} \]
• Repeat with sample of size 1,000,000
Matlab exercise: Uniform distribution

• N=10;
• Stats=1e3;
• r1=rand(Stats,1);
• r2=floor(N.*r1)+1;  % if need L:L+N, add L
• mean(r2)
• var(r2)
• [a,b]=hist(r2, 1:10);
• p_f=a./sum(a);
• figure; stem(b,p_f);
• Stats=1e6;  % repeat from line 3
Secretary problem or Picky bride problem

• A future bride has **known (large) number** \(- n \) – of suitors whom she dates one at a time
• She can easily **evaluate and rank the “quality” of suitors** relative to each other but has no idea of its distribution
• She has only **one chance to choose** her husband and cannot go back to the guys she dumped
• How can she **maximize the probability to choose the best groom** among \( n \) suitors?
Eugene Dynkin (1924 – 2014) solved the picky bride problem in 1963. He was a Russian and American mathematician, member of NAS. He has made contributions to the fields of probability and algebra. The Dynkin diagram, the Dynkin system, and Dynkin's lemma are all named after him.

Martin Gardner (1914 – 2010) described the secretary problem in Scientific American 1960. He was an American popular mathematics and popular science writer. Best known for “recreational mathematics”: He was behind the “Mathematical Games” section in Scientific American.
What should a poor bride do?

• She does not know the distribution of quality of suitors. Has to learn it on the fly
• Algorithm: look at the first $r$ suitors, remember the best among them
• Marry the first among next $n-r$ suitors who is better than the best among the first $r$ suitors
• How to choose $r$?
  • $r$ small – not enough information: the best among $r$ is not too good. You are likely to marry a looser.
  • $r=n-1$ – almost all information but you marry the last schmuck who is (likely not very good)
\[ P(r) = \sum_{i=1}^{n} P(\text{applicant } i \text{ is selected } \cap \text{ applicant } i \text{ is the best}) \]

\[ = \sum_{i=1}^{n} P(\text{applicant } i \text{ is selected} | \text{applicant } i \text{ is the best}) \times P(\text{applicant } i \text{ is the best}) \]

\[ = \left[ \sum_{i=1}^{r-1} 0 + \sum_{i=r}^{n} P \left( \text{the best of the first } i-1 \text{ applicants is in the first } r-1 \text{ applicants} \mid \text{applicant } i \text{ is the best} \right) \right] \times \frac{1}{n} \]

\[ = \sum_{i=r}^{n} \frac{r-1}{i-1} \times \frac{1}{n} = \frac{r-1}{n} \sum_{i=r}^{n} \frac{1}{i-1}. \]

Letting \( n \) tend to infinity, writing \( x \) as the limit of \( r/n \), using \( t \) for \( i/n \) and \( dt \) for \( 1/n \),

\[ P(x) = x \int_{x}^{1} \frac{1}{t} \, dt = -x \ln(x). \]

\[ \frac{dP(x)}{dx} = -\ln(x) - 1 \]

\(-\ln(x^*) - 1 = 0 \]

\[ x^* = 1/e = 0.3679 \]

Probability of picking the best is also \( 1/e = 0.3679 \)
Bonus matlab exercise: Picky bride

• Generate a **sample of size 100** of suitors with quality given by random number between 0 and 1
• Let bride select the first suitor exceeding the best suitor among the first
  – 10 suitors
  – 37 suitors
  – 90 suitors
• Compare to the best suitor among all 100
• Repeat for 1,000,000 brides and calculate the probability that she picked the best candidate in each case
• Test: for picking after dating 37 suitors the answer should be around 0.37