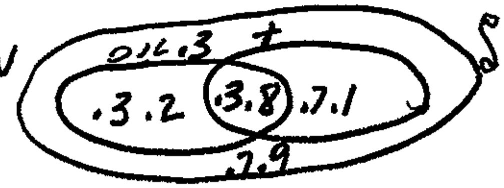
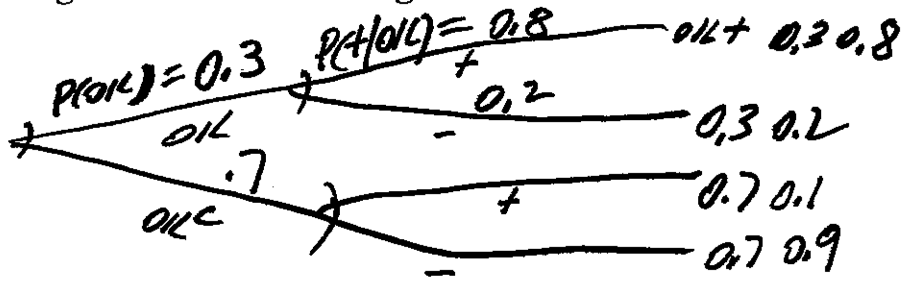
NOTE VENN



Key to prep for exam 1

1. Make a complete tree diagram for the following information:

P(OIL) = 0.3
P(+ | OIL) = 0.8
P(- | no OIL) = 0.9



Identify events like OIL+ in the tree and their probabilities (don't reduce).

2. From your tree in problem 1 determine

a. P(+ | no OIL) = **1 - P(- | no OIL)** = 1 − 0.9

b. P(-) = P(OIL -) + P(no OIL -)
= P(OIL) P(- | OIL) + P(no OIL) P(- | no OIL)
= .3 .2 + .7 .9 = .06 + .63 = .69

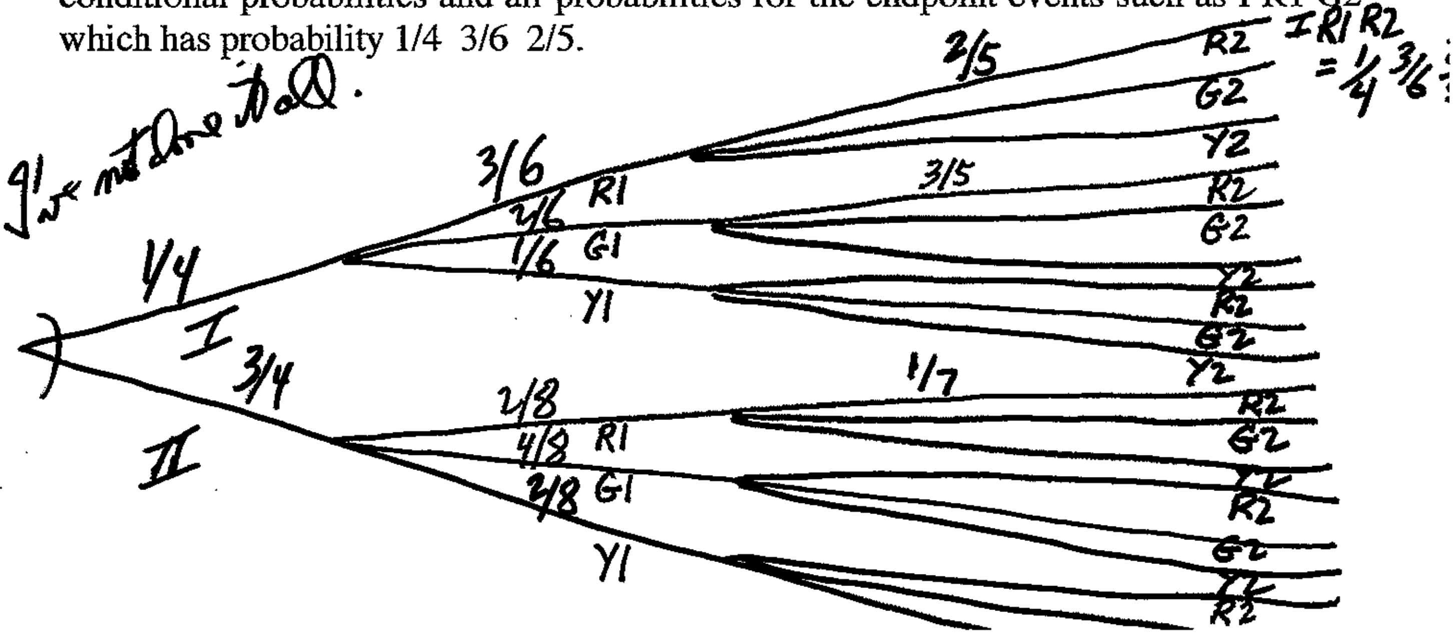
c. P(OIL | -). Is it less than P(OIL)? (make correction)
= P(OIL -) / P(-) = .3 .2 / (.3 .2 + .7 .9) = .08695

3. Box 1 will be chosen with probability 1/4. Otherwise box II will be chosen. We will then draw two balls from the chosen box, without replacement and with equal probability on those remaining.

Box I {R R R G G Y}
Box II {R R G G G G Y Y}

a. Make a complete tree diagram with root branch I, II and downstream branches for R1, G1, Y1, then subsequent downstream branches for R2, G2 Y2. Put in all of the conditional probabilities and all probabilities for the endpoint events such as I R1 G2 which has probability 1/4 3/6 2/5.

I've not done this.



b. Determine P(I R2).
= P(I) P(R2 | I)
= P(I) P(R1 | I) (order of the deal from box I does not matter)
= 1/4 3/6

c. Determine P(R2).
= P(I) P(R2 | I) + P(II) P(R2 | II)
= 1/4 3/6 + 3/4 2/8

d. Determine P(I | R2).
= P(I R2) / P(R2)
= (1/4 3/6) / (1/4 3/6 + 3/4 2/8)

4. If events A, B satisfy
P(A) = 0.5 P(B) = 0.3 P(B | A) = 0.2

a. P(AB) **= P(A) P(B | A) = 0.5 0.2 = 0.1 (notice, we don't use 0.3)**

b. $P(A \cup B)$ **= P(A) + P(B) - P(AB) = 0.5 + 0.3 - 0.5 0.2**

c. Are events A, B independent?
No, since 0.2 = P(B | A) ≠ P(B) = 0.3 as given.

d. Make a Venn diagram.
Key idea: P(A) = P(AB) + P(AB^C) and P(B) = P(AB) + P(A^CB).

AB	AB^C	A^CB	$A^C B^C$
0.1	**0.5 - 0.1 = 0.4**	**0.3 - 0.1 = 0.2**	**1 - rest = 0.3**

e. Determine P(neither A nor B) (i.e. $A^C \cap B^C$) as found in (d).
= 1 - (0.1 + 0.4 + 0.2) = 0.3

5. Engines A, B fail independently. Their respective failure probabilities are 0.001 and 0.002. Determine

a. P(both engines fail).
= (indep) 0.001 0.002 = 2 10^{-6}

b. P(at least one engine fails).
= (addn) 0.001 + 0.002 - 2 10^{-6} = 0.002998

c. P(neither fails).
= (indep) (1 - 0.001) (1 - 0.002) = 0.999 0.998 = 0.997002

6. Red and green dice are shaken in a cup and rolled. Let R be the number thrown by the red die and G be the number thrown by the green die. In the sample space of 36 outcomes calculate

a. $P(R = 5 \mid G = 2) = \frac{P(R=5 \text{ and } G=2)}{P(G=2)} = \frac{\#\text{favorable cases for } R=5 \text{ and } G=2\ /36}{\#\text{favorable cases for } G=2/36}$
= (1/36) / (6/36) = 1/6

b. Are the events R = 5, G = 2 independent?
Yes, since 1/6 = P(R = 5 | G = 2) = P(R = 5) = 6/36.

c. $P(R+2G < 8) = \frac{\#\text{favorable cases for } R+2G<8}{36}$ **= 9/36**

(total R + 2 G < 8 shown in each case)

	G 1	2	3	4	5	6
R 1	**3**	**5**	**7**			
2	**4**	**6**				
3	**5**	**7**				
4	**6**					
5	**7**					
6						

d. Is the event R+2G = 5 independent of the event R = 1? Check if
P(R=1 and R+2G < 8) = P(R=1) P(R+2G < 8).
P(R=1 and R+2G < 8) = P({R, G} are {1, 1} or {1, 2} or {1, 3}) = 3/36
P(R=1) P(R+2G < 8) = 6/36 9/36
No, the above two lines are not equal. So the events are dependent.

7. A random equal probability sample of n = 400 has
$\overline{x}$ = 12.6 s = 8.4

a. Margin of error for sample mean = **1.96 s / $\sqrt{n}$ = 1.96 8.4 / 20**

b. Around what percentage of the time will the population mean μ be covered by the sample interval $\overline{x}$ **± margin of error**?
around 95%

8. For the list of numbers {0, 4, 12} calculate

a. The sample mean $\overline{\boldsymbol{x}}$.
(0 + 4 + 12) / 3

b. The sample standard deviation **S** (remember, it has divisor n-1 inside the root).

$$\sqrt{\frac{(0-16/3)^2+(4-16/3)^2+(12-16/3)^2}{3-1}} = 6.1101$$

c. The margin of error of the sample mean.
$$1.96\, s / \sqrt{n} \;=\; 1.96\ 6.1101 \Big/ \sqrt{3} = 6.91423$$
Since n = 3 is so small the 95% interpretation does not apply.

9. A list has 100 ones and 200 zeroes. This could arise if we score each of 300 people 1 if they are insured and 0 if not. Calculate

a. The sample mean $\overline{\boldsymbol{x}}$ (it is necessarily equal to the relative proportion $\hat{p}$ of numbers 1 on the list).
$\hat{\boldsymbol{p}}$ **= 100 / 300 = 1/3**

b. The sample sd **S** of the list.
$$\sqrt{\frac{n}{n-1}}\ \sqrt{\hat{p}\left(1-\hat{p}\right)} = \sqrt{\frac{300}{299}}\ \sqrt{\frac{1}{3}\,\frac{2}{3}}$$

c. In statistical practice the margin of error for a 0-1 scored list is given as $\overline{x}\pm$ 1.96 $\sqrt{\frac{n-1}{n}}\,\frac{s}{\sqrt{n}}$ (which is the same as $\hat{p}$ ± 1.96 $\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$) rather than the usual form of margin of error $\overline{x}\pm$ 1.96 $\frac{s}{\sqrt{n}}$. For large n this is a minor numerical change in margin of error since $\sqrt{\frac{n-1}{n}}$ tends to limit 1. Calculate the margin of error for this 0-1 list.

$$1.96\,\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} = 1.96\,\frac{\sqrt{\frac{1}{3}\left(1-\frac{1}{3}\right)}}{\sqrt{n}}$$

d. If a population is sampled with replacement and equal probability and n is large with what (approximate) probability will the sample interval

$\overline{x} \pm$ margin of error

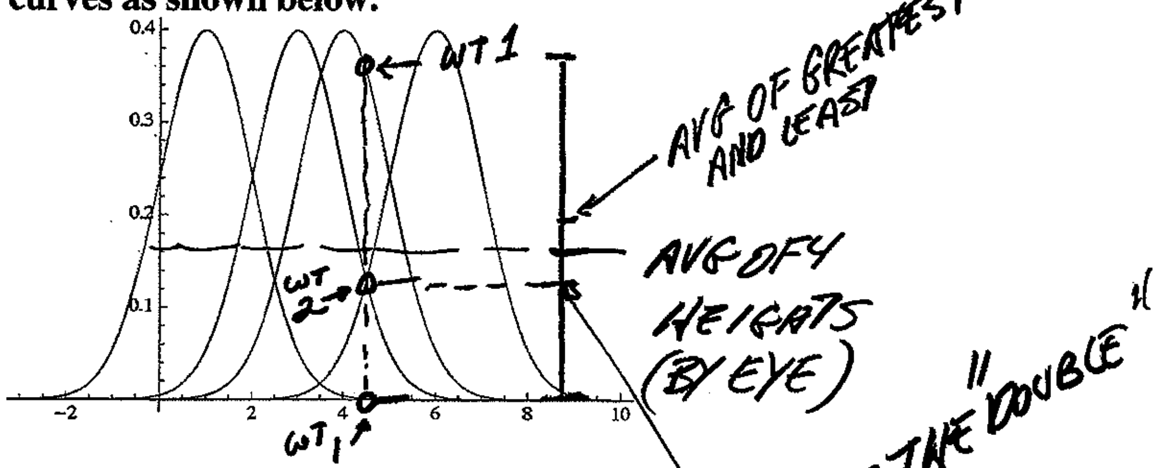
cover the population mean?

For n = 300 around 95%.

10. A list of 40 numbers has 6 in the interval [11, 12.5). What is the height of the probability histogram over the interval?

height = probability / width = (6 / 40) / (12.5 - 11) = 0.225

11. A probability density will be fit at x = 4.5 to four data points using the bell curves as shown below.



a. What are the four data values?

Below the tops of the bells {1, 3, 4, 6}.

b. What is the bandwidth being used (eyeball as best you can)?

around 1

c. Fit the density at x = 4.5.

See the figure.

12. Given a random variable X having the probability distribution below.

x	**-3**	**0**	**6**
p(x)	**0.1**	**0.7**	**0.2**

Determine

a. P(X < 5).
0.1 + 0.7 = 0.8

b. $P(X^2 < 10)$.
Collect the probabilities of x = -3 and 0 which are the only x whose square is less than 10.
ans. 0.1 + 0.7 = 0.8

c. E X.
-3 0.1 + 0 0.7 + 6 0.2 = 0.9

e. $E\,\frac{X}{1+X}$.
$$\frac{-3}{1-3}0.1 + \frac{0}{1+0}0.7 + \frac{6}{1+6}0.2 = 0.321429$$

f. Var X.
I'll use the short form.
$$E\,X^2 = \sum_x x^2 p(x) = (-3)^2\,0.1 + (0)^2\,0.7 + (6)^2\,0.2 = 8.1$$
$(E\,X)^2 = 0.9^2 = 0.81$ **(just a fluke it is so similar to above)**
Var X = 8.1 - 0.81 = 7.29

g. sd X.
$$\sqrt{\text{Var } X} = \sqrt{7.29} = 2.7$$

h. E (2X - 1) using rules of expectation.
2 E X - 1 = 2 0.9 - 1 = 0.8

i. Var (2X-1) using rules of variance.
= Var(2X) = 4 Var X = 4 7.29 = 29.16

j. sd(2X-1) using (i).
$$\sqrt{29.16} = 5.4$$

k. The distribution of Y = 2X-1 (list all distinct values y and their probabilities).

y	**2(-3)-1 = -7**	**2(0)-1 = -1**	**2(6)-1 = 11**
p(y)	**0.1**	**0.7**	**0.2**

l. Using direct calculation from (k) (not the rules) confirm your answer to (h).
$\sum_y y\, p(y) = (-7)\ 0.1\ +\ (-1)\ 0.7\ + (11)\ 0.2 = 0.8$ (same as (h))

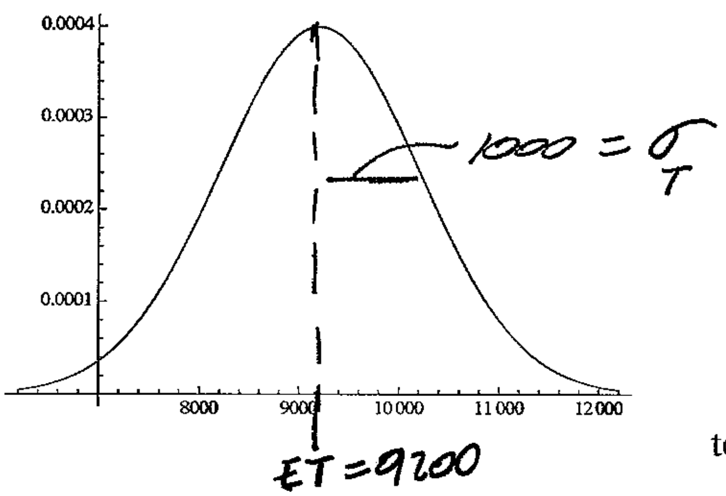
m. Using direct calculation from (k) (not the rules) confirm answer (i).
$E\,Y^2 = \sum_y y^2\, p(y) = (-7)^2\ 0.1\ +\ (-1)^2\ 0.7\ + (11)^2\ 0.2 = 29.8$
$\text{Var } Y = E\,Y^2 - (E\,Y)^2 = 29.8 - 0.8^2 = 29.16$ (same as (i)).

13. A venture has random return X with E X = 23 and sd X = 50. We contemplate 400 statistically independent ventures of the same kind. For the random total $T = X_1 + X_2 + \ldots + X_{400}$ of returns from 400 such ventures determine

a. E T (independence is **not** required for this part)
$E\,T = E\,(X_1 + X_2 + \ldots + X_{400}) = E\,X_1 + E\,X_2 + \ldots + E\,X_{400}$
$= 400\ E\,X = 400\ 23 = 9200$

b. sd **T (make correction)**
$\text{sd } T = \text{(indep) } \sqrt{400}\ \text{sd } X = 20\ 50 = 1000$

c. A sketch of the approximate distribution of T with (a), (b) shown as recognizable entities in your sketch.

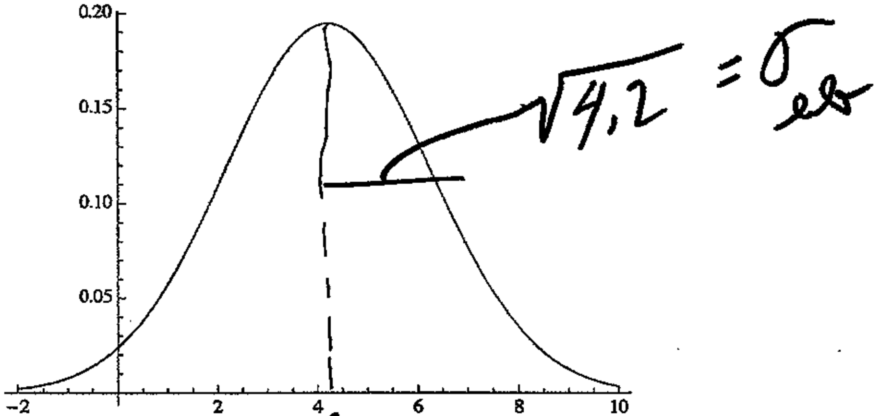


total t of 400

14. The number of error bits in a bit stream of a given length is thought to follow a Poisson distribution having the mean 4.2. Determine
a. The probability p(3) that a stream of the given length has exactly 3 error bits.

$$p(3) = e^{-4.2}\,\frac{4.2^3}{3!} = 0.185165$$

b. A sketch of the approximate distribution of the number of error bits in a stream of the given length. Label the mean and sd appropriately.



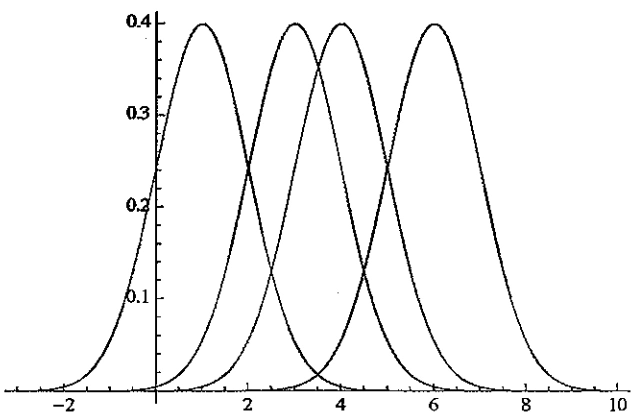
$E_{eb} = 4.2$ # error bits in a stream of a given length

c. The probability that more than 2 error bits occur in a stream of the given length.

$$1 - p(0) - p(1) - p(2) = 1 - \left(e^{-4.2}\,\frac{4.2^0}{0!} + e^{-4.2}\,\frac{4.2^1}{1!} + e^{-4.2}\,\frac{4.2^2}{2!}\right) = 0.78976$$

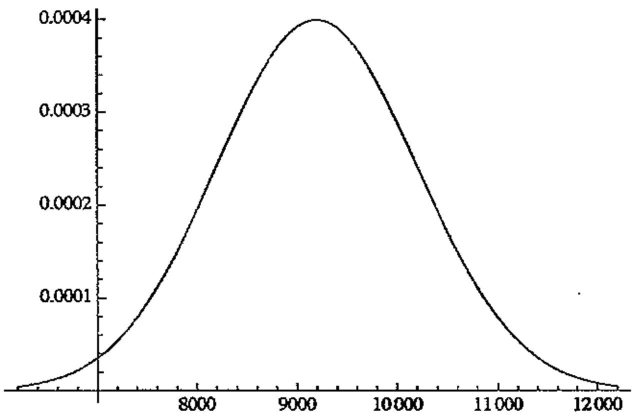
```
In[31]:= p[r_] := Exp[- r^2 / 2] / Sqrt[2 Pi]

         Plot[p[r - {1, 3, 4, 6}], {r, -3, 10}, PlotRange → All]
```



In[33]:= `Plot[Exp[- (r - 9200) ^2 / 2 000 000] / Sqrt[2 Pi 1 000 000], {r, 6200, 12 200}, PlotRange → All]`

Out[33]=



In[39]:= `Plot[Exp[- (r - 4.2) ^2 / (2 × 4.2)] / Sqrt[2 Pi 4.2],`
`{r, -2, 10}, PlotRange → All, AxesOrigin → {0, 0}]`

Out[39]=