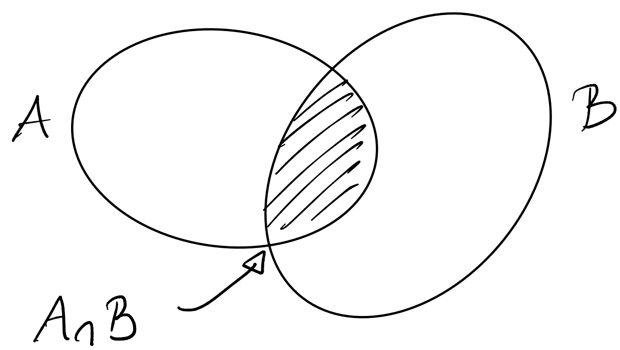
MATH 7241: Problems #1

Due date: Friday September 23

Reading: relevant background material for these problems can be found in the class notes, and in Ross (Sections 2.1 2.2, 2.3, 2.4) and in Grinstead and Snell (Chapters 1,2 3, 6).

Exercise 1 Let A and B be events such that $P(A) = 0.7$ and $P(B) = 0.9$. Find the largest and smallest possible values of $P(A \cup B) - P(A \cap B)$ (note: the event $A \cup B$ means either A or B or both are true, the event $A \cap B$ means both A and B are true).



$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B).$$

$$\Rightarrow \mathbb{P}(A \cup B) - \mathbb{P}(A \cap B) = 2\,\mathbb{P}(A \cup B) - \mathbb{P}(A) - \mathbb{P}(B)$$

$$= 2\,\mathbb{P}(A \cup B) - 1.6.$$

$\mathbb{P}(A \cup B)$: max. achieved when

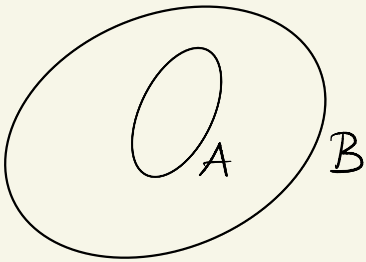
$A \cup B = S =$ whole sample space.

$$\Rightarrow \max(\mathbb{P}(A \cup B)) = 1$$

$\Rightarrow \max\left[P(A\cup B) - P(A\cap B)\right] = 2(1) - 1.6 = 0.4$

$P(A\cup B)$: min. achieved when

$A \subset B$



$\Rightarrow \min\left(P(A\cup B)\right) = P(B) = 0.9$

$\Rightarrow \min\left[P(A\cup B) - P(A\cap B)\right] = 2(0.9) - 1.6$

$= 0.2$

$\boxed{\textbf{Exercise 2}}$ Each of the following random variables is a well-known type. Identify each by name:
a) an airplane has four engines, and each engine may independently fail (very small probability!). X is the number of engines that fail.
b) flies are randomly landing on my pizza at a steady average rate. X is the number of flies that land on my pizza in the next five minutes.
c) a spammer sends a fake email to a new address every second. X is the number of attempts until somebody responds.
d) a farm raises several hundred thousand chickens. X is the weight of a randomly selected chicken.

a) X is binomial, with parameters

$n = 4$, p = prob. fail.

b) X is Poisson (number of arrivals)

c) X is geometric (number of trials until first success)

d) X is normal (random sample from large population)

Exercise 3 A town has five hotels; three people arrive and each randomly and independently selects a hotel. Find the probability that exactly two of them stay in the same hotel.

All choices are equally likely, so we can compute the probability by counting:

$$\mathbb{P}(\text{two in same hotel}) = \frac{\#\{\text{choices for two in same}\}}{\#\{\text{choices}\}}$$

It is easiest to count by assuming ordered events; so let's assume that the people arrive in order, one after another.

$$\Rightarrow \#\{\text{choices}\} = (5)(5)(5)$$

$\#\{$choices for two in same hotel$\}$

$= (5)(1)(4)$ ← first two are same, third is different

$+ (5)(4)(1)$ ← first & third same, second different

$+ (5)(4)(1)$ ← second and third same, first different

$= (5)(4)(3)$

$\Rightarrow \mathbb{P}($two in same hotel$)$

$$= \frac{(5)(4)(3)}{(5)(5)(5)} = \frac{12}{25}$$

Exercise 4 Find the mean of X, where the pdf is:

$$P(X = n) = (1-p)^2\, n\, p^{n-1}, \quad n = 1, 2, \ldots$$

[**Hint**: note that $\sum_{n=0}^{\infty} n(n-1)\, p^{n-2} = \frac{d^2}{dp^2} \sum_{n=0}^{\infty} p^n$].

Mean: $\mathbb{E}[X] = \sum_{n=1}^{\infty} n \; \mathbb{P}(X=n)$

(sum over all possible values)

(same function)

(pdf)

$$\Rightarrow \quad \mathbb{E}[X] = \sum_{n=1}^{\infty} n\,(1-p)^2\, n\, p^{n-1}$$

Here is the trick: recall geometric series

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \text{for } -1 < x < 1.$$

Take $\frac{d}{dx}$ of both sides:

$$\sum_{n=1}^{\infty} n\, x^{n-1} = \frac{1}{(1-x)^2}$$

Take $\frac{d}{dx}$ again:

$$\sum_{n=2}^{\infty} n(n-1)\, x^{n-2} = \frac{2}{(1-x)^3}$$

We want

$$\mathbb{E}[X] = (1-p)^2 \sum_{n=1}^{\infty} n^2 p^{n-1}$$

$$= (1-p)^2 \sum_{n=1}^{\infty} [n(n-1) + n] p^{n-1}$$

$$= (1-p)^2 p \sum_{n=2}^{\infty} n(n-1) p^{n-2}$$

$$+ (1-p)^2 \sum_{n=1}^{\infty} n p^{n-1}.$$

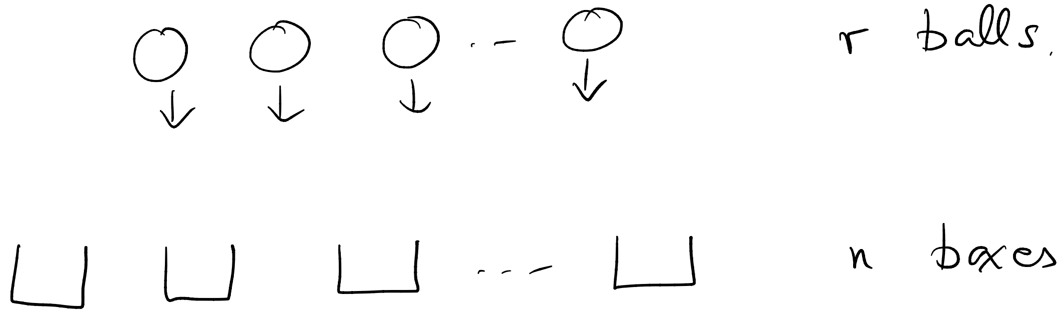
Now apply the formulas above with $x = p$:

$$\Rightarrow \mathbb{E}[X] = (1-p)^2 p \frac{2}{(1-p)^3} + (1-p)^2 \frac{1}{(1-p)^2}$$

$$= \frac{2p}{1-p} + 1$$

$$= \frac{1+p}{1-p}$$

Exercise 5 Randomly distribute r balls in n boxes. Find the probability that the first box is empty. Find the probability that the first two boxes are both empty.



$$P(\text{a given ball does not go in Box 1}) = 1 - \frac{1}{n}$$

$$\Rightarrow P(\text{Box \#1 is empty})$$

$$= P(\text{no balls go in Box \#1})$$

$$= P(\text{first ball not in Box \#1})$$
$$P(\text{second ball not in Box \#1})$$
$$\vdots$$
$$P(r^{\text{th}}\text{ ball not in Box \#1})$$

[we multiply these because the balls are dropped independently]

$$= \left(1 - \frac{1}{n}\right)^r$$

Furthermore,

$\mathbb{P}$(a given ball does not go in Box #1 or Box #2)

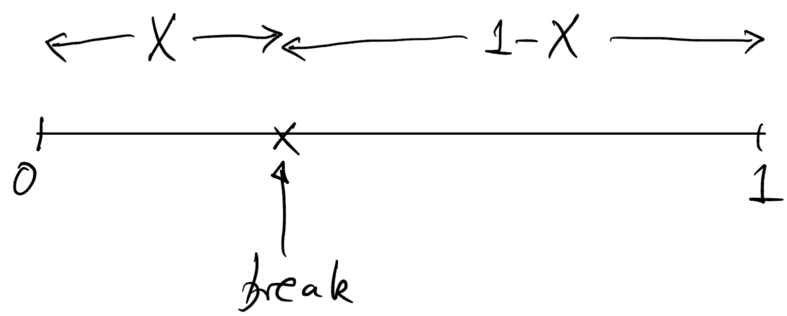
$$= \frac{n-2}{n} = 1 - \frac{2}{n}$$

So by the same reasoning,

$\mathbb{P}$(Boxes 1 and 2 are empty)

$$= \left(1 - \frac{2}{n}\right)^r$$

$\boxed{\textbf{Exercise 6}}$ We start with a stick of length 1, and break it in two pieces at a randomly chosen position (chosen uniformly over its length). Find the mean length of the longer end of the broken stick.



Stick is $[0,1]$; break it at position X,

where $X \sim U[0,1]$.

L = length of longer piece

$= \max(X, 1-X)$.

$$\Rightarrow \mathbb{E}[L] = \mathbb{E}[\max(X, 1-X)]$$

same function

$$= \int_0^1 \max(x, 1-x) \cdot 1 \cdot dx$$

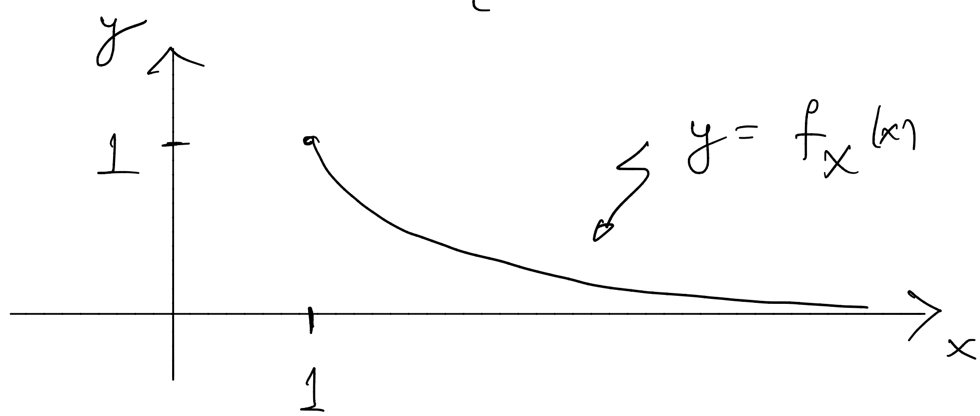
pdf of X

$$= \int_0^{\frac{1}{2}} (1-x)\, dx + \int_{\frac{1}{2}}^1 x \, dx$$

$$= \frac{3}{4}$$

Exercise 7 The current in a resistor is a random variable X. The pdf of X is $f(x) = e^{-(x-1)}$ for $x \geq 1$. The power dissipated in the resistor is $Y = X^2$. Find the pdf of Y.

pdf of X: $f_X(x) = \begin{cases} e^{-(x-1)} & x \geq 1 \\ 0 & \text{else} \end{cases}$



let $Y = X^2$. What is f_Y?

First find cdf of Y:

$$F_Y(y) = \mathbb{P}(Y \leq y)$$
$$= \mathbb{P}(X^2 \leq y)$$
$$= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\, dx$$

If $\quad y < 1 \Rightarrow \quad F_Y(y) = 0$

If $\quad y \geq 1$

$$\Rightarrow \quad F_Y(y) = \int_1^{\sqrt{y}} e^{-(x-1)}\,dx$$

$$= \int_0^{\sqrt{y}-1} e^{-u}\,du \qquad \begin{cases} u = x-1 \\ du = dx \end{cases}$$

$$= 1 - e^{-(\sqrt{y}-1)}$$

$$\Rightarrow \quad F_Y(y) = \begin{cases} 0 & y < 1 \\ 1 - e^{-(\sqrt{y}-1)} & y \geq 1 \end{cases}$$

$$\Rightarrow \quad f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \begin{cases} \frac{1}{2\sqrt{y}}\, e^{-(\sqrt{y}-1)} & y \geq 1 \\ 0 & \text{else} \end{cases}$$

Exercise 8 Derive the formula

$$\mathrm{VAR}[X_1 + X_2 + \cdots + X_n] = \sum_{k=1}^{n} \mathrm{VAR}[X_k] + 2\sum_{i<j} \mathrm{COV}(X_i, X_j) \tag{1}$$

Let $\mu_i = \mathbb{E}[X_i]$, $i = 1, \ldots, n.$

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mu_1 + \mu_2 + \cdots + \mu_n$$

$$\Rightarrow \mathrm{VAR}[X_1 + X_2 + \cdots + X_n]$$

$$= \mathbb{E}\left[\left(X_1 + X_2 + \cdots + X_n - (\mu_1 + \mu_2 + \cdots + \mu_n)\right)^2\right]$$

$$= \mathbb{E}\left[\left((X_1 - \mu_1) + (X_2 - \mu_2) + \cdots + (X_n - \mu_n)\right)^2\right]$$

$$= \mathbb{E}\left[(X_1 - \mu_1)^2\right] + \mathbb{E}\left[(X_2 - \mu_2)^2\right] + \cdots$$

$$+ 2\mathbb{E}\left[(X_1 - \mu_1)(X_2 - \mu_2)\right]$$

$$+ 2\mathbb{E}\left[(X_1 - \mu_1)(X_3 - \mu_3)\right] + \cdots$$

$$+ 2\mathbb{E}\left[(X_{n-1} - \mu_{n-1})(X_n - \mu_n)\right]$$

$$= \mathrm{VAR}[X_1] + \mathrm{VAR}[X_2] + \cdots + \mathrm{VAR}[X_n]$$

$$+ 2\,\mathrm{COV}[X_1, X_2] + 2\,\mathrm{COV}[X_1, X_3]$$

$$+ \cdots + 2\,\mathrm{COV}[X_{n-1}, X_n]$$

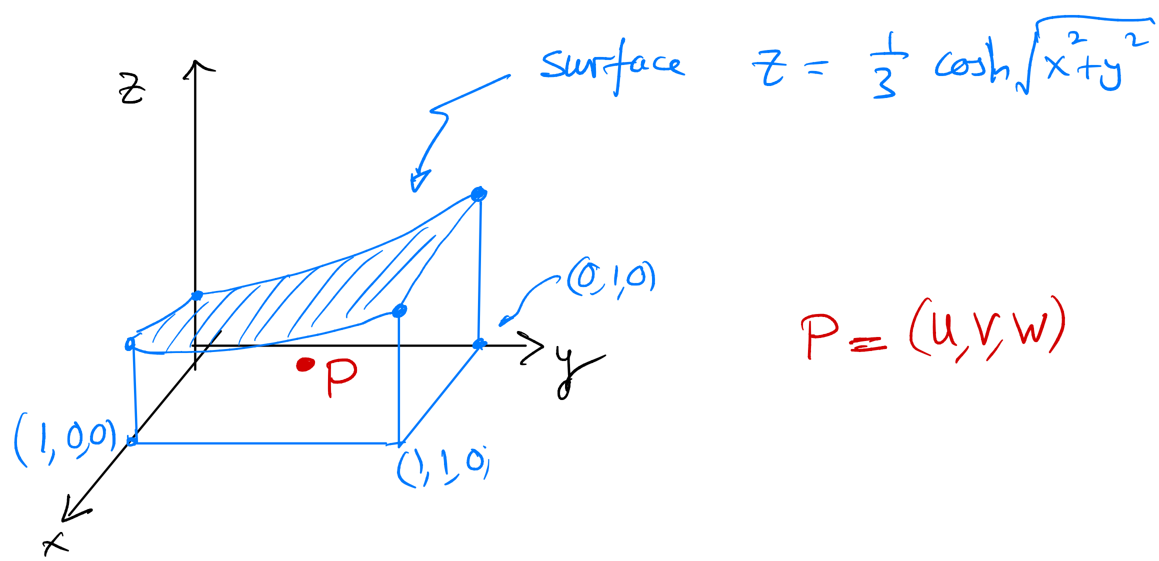
$$= \sum_{i=1}^{n} \mathrm{VAR}[X_i]$$

$$+ 2 \sum_{1 \le i < j \le n} \mathrm{COV}[X_i, X_j]$$

$\boxed{\textbf{Exercise 9}}$ Find a random number generator that generates uniformly on $[0,1]$ (for example the command *rand* in Matlab). Using this generator, estimate the volume of the region under the surface

$$z = \frac{1}{3}\cosh\sqrt{x^2+y^2}$$

and above the unit square $0 \le x \le 1$, $0 \le y \le 1$. [Note: generate three independent uniform random variables for each run, corresponding to the three coordinates of a random point in the unit cube. Do enough runs to be confident that you have an accurate estimate of the first two decimal places].



Let $\{U,V,W\}$ be independent uniform r.v.'s in $[0,1]$. Then $P=(U,V,W)$ is a randomly chosen point in the unit cube $[0,1]^3$. Therefore

$$\mathbb{P}(\text{ P lies below surface}) = \frac{\text{volume of region below surface}}{\text{volume of unit cube}}$$

= volume of region below surface

$$= \mathbb{P}\left(W \leq \frac{1}{3} \cosh \sqrt{U^2+V^2} \right).$$

To estimate this probability, generate N independent triples (U_i, V_i, W_i), $i=1,\dots,N$, and define

$$R_N = \frac{\#\left\{ i : W_i \leq \frac{1}{3} \cosh \sqrt{U_i^2+V_i^2} \right\}}{N}.$$

As $N \to \infty$, the Law of Large Numbers implies that R_N converges to the probability that we want.

Here are results for several values of N (using Matlab).

N	Trial 1	Trial 2	Trial 3	Time per trial (sec)
1000	0.439	0.438	0.449	0.001
10,000	0.453	0.450	0.457	0.003
10^5	0.455	0.452	0.453	0.01
10^6	0.4531	0.4534	0.4534	0.15

Actual value of integral is

$$\int_0^1 \int_0^1 \frac{1}{3} \cosh \sqrt{x^2+y^2} \, dx \, dy = 0.4534.$$

$\boxed{\textbf{Exercise 11}}$ In class we considered this problem: "An urn contains n Red balls and m Black balls. Suppose that k balls are withdrawn from the urn, and let X be the number of Red balls among these. Find $\mathbb{E}[X]$ assuming (i) replacement, and (ii) no replacement." Using the same reasoning as in class, compute VAR[X] assuming (i) replacement, and (ii) no replacement. [Hint: use the formula from Exercise 8 above. The answers will be different for the two cases].

$$X = R_1 + R_2 + \cdots + R_k$$

i) With replacement the $\{R_i\}$ are IID.

$$\Rightarrow VAR[X] = \sum_{i=1}^{k} VAR[R_i]$$

$$= k\, VAR[R_1].$$

R_1	0	1
prob	$\frac{m}{n+m}$	$\frac{n}{n+m}$

$$\Rightarrow VAR[R_1] = \frac{nm}{(n+m)^2}$$

$$\Rightarrow VAR[X] = \frac{knm}{(n+m)^2}$$

ii) Without replacement, the $\{R_i\}$ are not independent. (Assume that $k \le n+m$).

$$VAR[X] = \sum_{i=1}^{k} VAR[R_i] + 2 \sum_{1 \le i < j \le k} COV[R_i, R_j]$$

As we showed in class, the pdf of R_i is the same for all $i = 1, \dots, k$

$$\Rightarrow VAR[R_i] = VAR[R_1] = \frac{nm}{(n+m)^2}$$

Similarly $COV[R_i, R_j]$ is the same for all pairs (i,j). Note that

$$COV[R_i, R_j] = \mathbb{E}[R_i R_j] - \mathbb{E}[R_i]\,\mathbb{E}[R_j]$$

Now $\mathbb{E}[R_i R_j] = \mathbb{P}(R_i = 1, R_j = 1)$.

As we showed in class, the probability of any sequence of colors

e.g. $R_1 B_2 B_3 R_4 R_5$

is the same if we permute it

e.g. $B_1 B_2 R_3 R_4 R_5$

so for example

$$\mathbb{P}(R_1=1, R_3=1) = \mathbb{P}(R_1=1, R_2=1, R_3=1) + \mathbb{P}(R_1=1, R_2=0, R_3=1)$$

$$= \mathbb{P}(R_1=1, R_2=1, R_3=1) + \mathbb{P}(R_1=1, R_2=1, R_3=0)$$

$$= \mathbb{P}(R_1=1, R_2=1).$$

Similarly $\mathbb{P}(R_i=1, R_j=1) = \mathbb{P}(R_1=1, R_2=1)$ for all $i<j\le k$

Now joint pdf:

$R_2 \backslash R_1$	0	1
0	$\dfrac{m(m-1)}{(n+m)(n+m-1)}$	$\dfrac{mn}{(n+m)(n+m-1)}$
1	$\dfrac{mn}{(n+m)(n+m-1)}$	$\dfrac{n(n-1)}{(n+m)(n+m-1)}$

$$\Rightarrow \mathbb{E}[R_1R_2] = \mathbb{P}(R_1=1, R_2=1) = \frac{n(n-1)}{(n+m)(n+m-1)}$$

$$\Rightarrow \quad COV[R_1, R_2] = \frac{n(n-1)}{(n+m)(n+m-1)} - \left(\frac{n}{n+m}\right)^2$$

$$= \frac{-nm}{(n+m)^2(n+m-1)}$$

$$\Rightarrow \quad VAR[X] = k\frac{nm}{(n+m)^2} - k(k-1)\frac{nm}{(n+m)^2(n+m-1)}$$

Exercise 11 A typing firm has three typists A,B and C. The number of errors per 100 pages made by typist A is a Poisson random variable with mean 2.6; the number of errors per 100 pages made by typist B is a Poisson random variable with mean 3; the number of errors per 100 pages made by typist C is a Poisson random variable with mean 3.4. A manuscript of 300 pages is sent to the firm. Let X denote the number of errors in the typed manuscript.
a) Assume that one typist is randomly selected to do all the work. Find the mean and variance of X.
b) Assume instead that the work is divided into three equal parts which are given to the three typists. Find the mean and variance of X in this case.

a) $$\mathbb{E}[X] = \mathbb{E}[X|A]\tfrac{1}{3} + \mathbb{E}[X|B]\tfrac{1}{3} + \mathbb{E}[X|C]\tfrac{1}{3}$$

$$\mathbb{E}[X^2] = \mathbb{E}[X^2|A]\tfrac{1}{3} + \mathbb{E}[X^2|B]\tfrac{1}{3} + \mathbb{E}[X^2|C]\tfrac{1}{3}$$

Typist A: Poisson, mean $= 3(2.6) = 7.8 = \lambda_A$

$$\Rightarrow \mathbb{E}[X|A] = \lambda_A$$

$$\mathrm{VAR}[X|A] = \lambda_A$$

$$\Rightarrow \mathbb{E}[X^2|A] = \mathrm{VAR}[X|A] + (\mathbb{E}[X|A])^2$$

$$= \lambda_A + \lambda_A^2.$$

Similarly for B, C with $\lambda_B = 9$, $\lambda_C = 10.2$

$$\Rightarrow \mathbb{E}[X] = \tfrac{1}{3}(\lambda_A + \lambda_B + \lambda_C) = 9$$

$$\mathbb{E}[X^2] = \tfrac{1}{3}(\lambda_A + \lambda_B + \lambda_C + \lambda_A^2 + \lambda_B^2 + \lambda_C^2) = 90.96$$

$$\Rightarrow VAR[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 9.96$$

b) $X = X_A + X_B + X_C$

independant Poisson, rates $2.6, 3, 3.4$

$$\Rightarrow \mathbb{E}[X] = \mathbb{E}[X_A] + \mathbb{E}[X_B] + \mathbb{E}[X_C] = 9$$

$$VAR[X] = VAR[X_A] + VAR[X_B] + VAR[X_C]$$

$$= 9.$$