1) The probability density of X is given by

$$f(x) = \begin{cases} \sqrt{\frac{2}{\pi}}\,\frac{x^2 e^{\frac{-x^2}{2\sigma^2}}}{\sigma^3} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where $\sigma > 0$. Calculate $E(X)$ (10 points).

First, note that

$$E(X) = \int_0^\infty \sqrt{\frac{2}{\pi}}\,\frac{x^3 e^{\frac{-x^2}{2\sigma^2}}}{\sigma^3}\,dx.$$

Use the equation

$$\Gamma(\alpha) = 2^{1-\alpha}\int_0^\infty z^{2\alpha-1}e^{-z^2/2}\,dz$$

to evaluate this integral. Make the substitution $z = \frac{x}{\sigma}$ and let $\alpha = 2$. So

$$\Gamma(2) = \frac{1}{2}\int_0^\infty \frac{x^3}{\sigma^3}\,e^{\frac{-x^2}{2\sigma^2}}\,\frac{dx}{\sigma}.$$

Now use the fact that $\Gamma(2) = 1$ to obtain

$$2\sigma = \int_0^\infty \frac{x^3}{\sigma^3}\,e^{\frac{-x^2}{2\sigma^2}}\,dx.$$

This implies that

$$2\sigma\sqrt{\frac{2}{\pi}} = \int_0^\infty \sqrt{\frac{2}{\pi}}\,\frac{x^3 e^{\frac{-x^2}{2\sigma^2}}}{\sigma^3}\,dx.$$

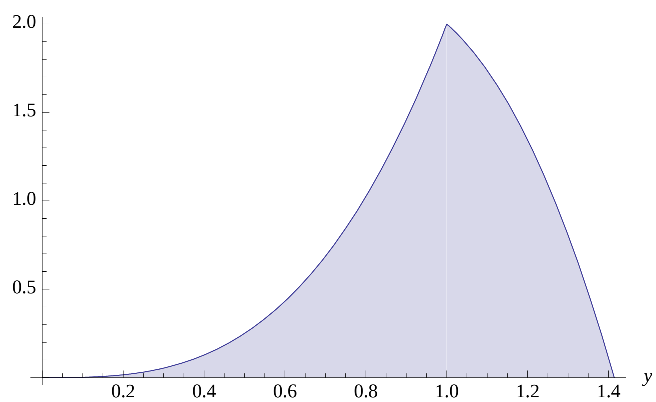
2) Let X_1 and X_2 be independent random variables such that both X_1 and X_2 have the uniform density on the interval $(0,1)$. Find the probability density of $Y = \sqrt{X_1 + X_2}$ (10 points).

We first consider the transformation $X = X_1 + X_2$. The distribution of X is the well-known tent distribution with density function

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2 - x & \text{for } 1 \le x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

Now the required transformation is $Y = \sqrt{X}$. Here $w(y) = y^2$, and so, $w'(y) = 2y$. The density function of Y is $f(w(y)) \cdot |w'(y)|$, that is,

$$g(y) = \begin{cases} 2y^3 & \text{for } 0 < y < 1 \\ 2y(2 - y^2) & \text{for } 1 \le y < \sqrt{2} \\ 0 & \text{elsewhere} \end{cases}$$



3) The waiting time at your doctor's office is exponentially distributed with average waiting time of 14 minutes. Calculate the probability that the minimum waiting time of your next seven visits will be less than 2 minutes. Assume that the waiting times on all your visits are independent of each other (10 points).

The minimum of 7 independent exponential random variables with mean 14 minutes is an exponential random variable with mean 2 minutes. Now the probability that the minimum waiting time is less than 2 minutes is

$$\int_0^2 \frac{1}{2} e^{-\frac{x}{2}} \, dx = 1 - \frac{1}{e} \approx 0.632120559.$$

4) There are an unknown amount of marbles in a black jar. You know that exactly 5 marbles are red. Someone else randomly picks out 6 marbles from the jar and you see that 3 of these selected marbles are red. Use the method of maximum likelihood to estimate the number of marbles in the jar (10 points).

Let N be the unknown amount of marbles and let E be the event that 6 marbles are randomly chosen from N marbles and exactly 3 of these are red. Below, we compute the probability of E for $N = 8$, 9, 10, 11, 12, and 13.

N	$P(E)$
8	0.357143
9	**0.47619**
10	**0.47619**
11	0.4329
12	0.378788
13	0.32634

The general formula used above is

$$P(E) = \frac{\binom{5}{3}\binom{N-5}{3}}{\binom{N}{6}}.$$

The event E is most likely when N is 9 or 10. These are the maximum likelihood estimators for the number of marbles.

5) A pollster wishes to estimate the percentage of voters who favor a certain candidate. She wishes to be sure with probability 0.99 that the error in the resulting estimate will not exceed 3%. How many voters should she interview? (10 points)

We use the maximum possible standard deviation of a half to consider the error in the worst-case scenario. In this situation, we require the error to be less than 0.03, that is,

$$2.58\frac{1}{2\sqrt{n}} \leq 0.03$$

So

$$n \geq \frac{(2.58)^2}{4(0.03)^2} = 1849.$$

This means that at least 1849 people should be polled.

6) Candidates Oldwhiteguy and Soccermom are fighting for the Independent votes in an election. For the next three months they plan to run daily spots targeting these Independent voters. The networks require these commercials and the schedule when to run these ads in advance, and each candidate is unaware of the other candidate's schedule. Collected data from polling indicates the following trends. When they both run negative ads on the same day, candidate Soccermom wins 150 of the Independent voters over to her side. When they both run positive ads, candidate Soccermom wins 50 Independent voters over. When some candidate runs a negative ad and the other a positive ad, candidate Oldwhiteguy gains 100 of the Independent voters. The votes won over are "solid votes", that is, once the Independent voter aligns himself/herself with a candidate, he/she will stick with that candidate. What is candidate Soccermom's optimal strategy? On average, how many Independent votes does Soccermom win/lose daily while following her optimal strategy? (10 points)

		OWG +	OWG −
Mom	+	50	−100
Mom	−	−100	150

Let x be the probability with which Soccermom goes positive. The expected number of votes gained are

$$\begin{aligned} 50x - 100(1-x) &= 150x - 100 && \text{when OWG runs a + add} \\ -100x + 150(1-x) &= -250x + 150 && \text{when OWG runs a − add} \end{aligned}$$

Equate these two expected values and solve for x in order to obtain $x = \frac{5}{8}$ and $1-x = \frac{3}{8}$. So candidate Soccermom's optimal strategy is to go positive with probability $\frac{5}{8}$ and attack with negative adds with probability $\frac{3}{8}$.

When Soccermom follows this strategy, she can hold her expected losses to as low as 6.25 voters per day.

Note that since the payoff matrix is symmetric about the diagonal, candidate Oldwhiteguy has the same optimal strategy.

1) A die is rolled 16 times. It shows up 6 four times, 5 once, 4 twice, 3 four times, 2 once, and 1 four times. Is this die balanced? (5 points)

We will test the hypothesis that *the die is balanced*, using a 95% confidence interval. The number of 6's, 3's, and 1's seem high. If the die was perfectly balanced, we would see 6's, 3's, and 1's half the time (8 out of 16 times). However, we see these numbers a total of 12 times. Calculate the z-score with standard deviation of a half. Sixteen is large enough to apply the Central Limit Theorem for the Binomial distribution with $\theta = \frac{1}{2}$.

$$z = \frac{\frac{12}{16} - \frac{1}{2}}{\frac{1}{2\sqrt{16}}} = 2.$$

This lies outside the 95% confidence interval. We claim that the the die is not balanced based on this test. Note that if we use a 99% confidence interval, then we cannot reject the hypothesis that the die is balanced. It may be argued that since the sample size is just 16 (less than 30), maybe a 99% confidence interval is more appropriate.

2) A random-number machine outputs a random integer (not necessarily positive) every minute from an unknown but fixed range of integers. You know that every integer from this unknown range is equally likely. The last 25 integers from the machine are -10, 5, 9, 3, 18, 5, -11, -10, 16, 11, -1, -20, 16, -11, 5, -4, -12, 8, -5, -6, 2, 11, 14, 20, and -2. Rank the following events from most likely to least likely (5 points).

A: The next integer will be greater than 20.
B: The next integer will be less than -15
C: The next integer will be in between -5 and 5.
D: The next integer will be even.

We first make the simplifying assumption that we are dealing with a uniform continuous distribution. The endpoints, α and β, are unknown quantities here. One way to estimate α and β is to "guess" that the sample mean is equal to the mean and the sample variance is equal to the variance of the distribution. The average of the the 25 numbers is $\frac{51}{25}$, and so, let

$$\frac{\beta + \alpha}{2} = \frac{51}{25}. \tag{1}$$

The sample variance of the 25 numbers is $\frac{35887}{300}$, and so, let

$$\frac{(\beta - \alpha)^2}{12} = \frac{35887}{300}. \tag{2}$$

These equations can be easily solved to see that $\alpha = -17.92$ and $\beta = 19.96$. The estimate of α doesn't quite correspond with our observations since -20 is one of the last 25 numbers. So we estimate that $\alpha = -20$ and $\beta = 20$. Now we can calculate the probabilities of all the events. Event D is the most likely with probability close to a half. The probability of A is 0, the probability of B is $\frac{5}{41}$, and the probability of C is $\frac{9}{41}$. Based on these estimates, the list of events from most likely to least likely is D, C, B, and A.

3) You start with a \$100 at a casino. This casino is running a Christmas Special and they have a fair game—there is a 50% chance of winning what you bet and a 50% chance of losing it. You decide to play the minimum bet of \$5 each time. What is the probability of increasing your cash to \$250 before going broke with this strategy? (5 points)

Let p be the probability of reaching \$250 before going broke. Since this is a fair game, if you bet the same amount every time, your expected earnings should be zero. In other words, the expected value of your cash should be \$100. So

$$100 = p\,250 + (1 - p)0.$$

It now follows that $p = \frac{2}{5}$.

If the probability of winning a bet is reduced to 48% (like in Blackjack), then the probability of you getting to \$250 before going bust is approximately 7%—it is really this low! Under fair play, the probability of you doubling your money before going broke is obviously a half. If the game is 52% in favor of the house, like Blackjack is, then the probability of you reaching \$200 before going broke is around 17%. So betting the same amount every time is not a good strategy. You need a better system.