1. Let A, B and C are three events such that P(A) = 0.45, P(B) = 0.30, P(C) = 0.35, $P(A \cup B) = 0.60, P(A \cup C) = 0.60, P(B \cup C) = 0.50, P(A \cup B \cup C) = 0.70$.

(a) Compute $P(A \cap B), P(A \cap C), P(B \cap C)$.

(b) Compute $P(A \cap B \cap C)$.

(c) Compute the probability that exactly one of A, B and C happens.

Solution. (a) $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.15$, $P(A \cap C) = P(A) + P(C) - P(A \cup C) = 0.20$, $P(B \cap C) = P(B) + P(C) - P(B \cup C) = 0.15$.

(b) Since

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

we have

$$P(A \cap B \cap C) = P(A \cup B \cup C) - P(A) - P(B) - P(C) + P(A \cap B) + P(A \cap C) + P(B \cap C) = 0.10.$$

(c) We have $P(A \cap B \cap C') = P(A \cap B) - P(A \cap B \cap C) = 0.05$, $P(A \cap B' \cap C) = P(A \cap C) - P(A \cap B \cap C) = 0.10$, $P(A' \cap B \cap C) = P(B \cap C) - P(A \cap B \cap C) = 0.05$. Now

$$P(A \cap B' \cap C') = P(A) - P(A \cap B \cap C') - P(A \cap B' \cap C) - P(A \cap B \cap C) = 0.25.$$

$$P(A' \cap B \cap C') = P(B) - P(A \cap B \cap C') - P(A' \cap B \cap C') - P(A \cap B \cap C) = 0.10.$$

$$P(A' \cap B' \cap C) = P(C) - P(A' \cap B \cap C) - P(A \cap B' \cap C) - P(A \cap B \cap C) = 0.05.$$

So the answer is $0.25 + 0.10 + 0.05 = 0.40$.

2. You are dealt a 2-card hand from a well-shuffled deck of 52.

(a) What is the probability of the union of the event that both cards are aces with the event that both cards are red(hearts or diamons)?

(b) That is the probability that both cards have the same denomination (e.g. both are 2s or both are jacks)?

(c) What is the conditional probability that the second card is a picture card (J, Q, or K) given that at least one of the two cards is a picture card ?

Solution. (a) Let A be the event that both cards are aces, and R be the event that both cards are red. Then $P(A) = \dfrac{\binom{4}{2}}{\binom{52}{2}}$ since there are 4 aces. Likewise $P(R) = \dfrac{\binom{26}{2}}{\binom{52}{2}}$. Next

$P(AR) = \dfrac{1}{\binom{52}{2}}$ since there is only one good combination $(\diamondsuit A, \heartsuit A)$. Hence the answer is

$$\frac{\binom{4}{2} + \binom{26}{2} - 1}{\binom{52}{2}}.$$

(b) After the first card is drawn there are 51 cards left in the desk and 3 cards have the same denomination as the first card. Hence the answer is $\frac{3}{51}$.

(c) Let B_1, B_2 be the events that the first (respectively, second) card is a picture card. There are $12(= 3 \times 4)$ picture cards. Hence

$$P(B_1) = P(B_2) = \frac{12}{52} \approx 0.23 \text{ and } P(B_1B_2) = \frac{\binom{12}{2}}{\binom{52}{2}} \approx 0.05.$$

Thus if B is an event that at least one card is a picture card, then

$$P(B) = P(B_1) + P(B_2) - P(B_1 \cap B_2) \approx 0.41.$$

Hence $P(B_1|B) = \frac{P(B_1 \cap B)}{P(B)} = \frac{P(B_1)}{P(B)} \approx 0.56.$

3. In a certain city there are two food stores. 70% of the the population use *Cheap FoodStore* and 30% use *Green FoodStore*. Among the shoopers of *Cheap FoodStore* 70% have household income of less than 50000 per year while among the shoopers of *Green FoodStore* 50% have household income of less than 50000 per year.

(a) Find the probability that a randomly chosen citizen uses *Green FoodStore* and has household income 50000 or more per year.

(b) What proportion of the citizens have household income 50000 or more per year.

(c) Given that a person has a household income 50000 or more per year how likely he is to shop at *Green FoodStore*.

Solution. Let C be event that a citizen uses *Cheap FoodStore* and G be event that a citizen uses *Green FoodStore*. Let R be the event that a citizen has a household income of 50000 or more per year. Note that $P(R|C) = 1 - P(R'|C) = 0.3$ and $P(R|G) = 1 - P(R'G) = 0.5$.

(a) $P(RG) = P(G)P(R|G) = 0.3 * 0.5 = 0.15$.

(b) $P(R) = P(RG) + P(RC) = 0.3 * 0.5 + 0.7 * 0.3 = 0.36$.

(c) $P(G|R) = P(RG)/P(R) = 5/12$.

4. Let X have cumulative distribution function $F_X(x) = x^2$ for $0 \leq x \leq 1$ (and $F_X(x) = 0$ for $x < 0$ and $= 1$ for $x > 1$),

(a) Find the density of $f_X(x)$;
(b) find the probability density function $f_V(v)$ of $V = 3X+1$;
(c) Compute the median, 25th and 75th percentiles of X.

Solution. (a) $F_X(x) = (x^2)' = 2x$ if $x \in [0,1]$ (and 0 otherwise).
(b) V takes value between $3*0+1 = 1$ and $3*1+1 = 4$. If $v \in [1,4]$ then

$$P(V < v) = P(3X+1 < v) = P(X < \frac{v-1}{3}) = \left(\frac{v-1}{3}\right)^2.$$

Thus

$$f_V(v) = \frac{dv}{dv}\left[\left(\frac{v-1}{3}\right)^2\right] = 2 \times \frac{v-1}{3} \times \frac{1}{3} = \frac{2v-2}{9}.$$

(c) Equation for the k-th percentile is $x^2 = \frac{k}{100}$ so that $x_k = \sqrt{\frac{k}{100}}$. Thus $m = \frac{\sqrt{2}}{2}$, $x_{75} = \frac{\sqrt{3}}{2}$, $x_{25} = \frac{1}{2}$.

5. The continuous random variable Y has density $f_Y(y) = \frac{2}{27}y(y+1)$ for $0 \le y \le 3$ (and $f_Y(y) = 0$ otherwise).
(a) Find the cumulative distribution function of Y.
(b) Find $E(1/Y)$.
(c) If $Y_1, Y_2 \dots Y_{10}$ are independent random variables with distribution Y find the distribution of $M = \max(Y_1, Y_2 \dots Y_{10})$.

Solution. (a) $f_Y(y) = \frac{2}{27}(y^2+y)$. Thus

$$F_Y(y) = \frac{2}{27}\int_0^y (s^2+s)ds = \frac{2}{81}y^3 + \frac{1}{27}y^2.$$

$$(b) \quad EY = \frac{2}{27}\int_0^3 \frac{y^2+y}{y}dy = \frac{2}{27}\int_0^3 (y+1)dy = \frac{y^2+2}{27}\Big|_0^3 = \frac{11}{27}.$$

$$(c) \quad P(M < m) = P(Y_1 < m \dots Y_{10} < m) = P(Y < m)^{10} = \left[\frac{2}{81}m^3 + \frac{1}{27}m^2\right]^{10}.$$

6. 1000 independent random variables W_j, $j = 1, \dots, 1000$ have been simulated, according to an exponential distribution with parameter $\lambda = 1$.
(a) Let N be the number among the variables W_j which are larger than $\ln(500)$. What is the exact probability distribution of N?
(b) What is the approximate probability that N is 3 or less?
(c) Answer questions (a), (b) if the number $\ln(500)$ is replaced by $\ln(5)$.

Solution. (a) $P(W > \ln 500) = e^{-\ln 500} = \frac{1}{500}$. Accordingly $N \sim$Bin$(1000, \frac{1}{500})$.

(b) Since $1000 \gg 1$ and $1000 \times \frac{1}{500} = 2$ is not too large we have $N \approx$Pois(2). Thus

$$P(N \leq 3) = P(N=0)+P(N=1)+P(N=2)+P(N=3) \approx e^{-2}\left(\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}\right) \approx 0.86.$$

(c) $P(W > \ln 5) = e^{-\ln 5} = \frac{1}{5}$. Accordingly $N \sim$Bin$(1000, \frac{1}{5})$.

$$P(N \leq 3) = P(N=0) + P(N=1) + P(N=2) + P(N=3)$$

$$= \left(1 \times \left(\frac{4}{5}\right)^{1000} + \frac{1000}{1!}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^{999} + \frac{1000*999}{2!}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^{998} + \frac{1000*999*998}{6!}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^{997}\right).$$

Since the last term is much larger than all other terms we get

$$P(N \leq 3) \approx \frac{1000^3}{6}\frac{4^{997}}{5^{1000}} \approx 3.2 \times 10^{-91}.$$

7. Let (X, Y) have joint density $f_{X,Y}(x, y) = 2xy + 8x^3y^3$ if $0 \leq x \leq 1$ and $0 \leq y \leq 1$ (and $f_{X,Y}(x, y) = 0$ otherwise).

(a) Compute the marginal density of X.

(b) Compute EX and VX.

(c) Find the $Cov(X, Y)$.

Solution.

$$(a) \quad p_X(x) = \int_0^1 2xydy + \int_0^1 8x^3y^3dy = x + 2x^3.$$

$$(b) \quad EX = \int_0^1 (x+2x^3)xdx = \int_0^1 x^2dx + \int_0^1 2x^4dx = \frac{1}{3} + \frac{2}{5} = \frac{11}{15}.$$

Likewise $EY = \frac{11}{15}$.

$$EX^2 = \int_0^1 (x+2x^3)x^2dx = \int_0^1 x^3dx + \int_0^1 2x^5dx = \frac{1}{4} + \frac{2}{6} = \frac{7}{12}.$$

Hence $VX = \frac{7}{12} - \frac{11^2}{15^2} = \frac{7*75-121*4}{900} = \frac{41}{900}$.

$$(c) \quad E(XY) = \int_0^1\int_0^1 (2xy + 8x^3y^3)xydxdy = \int_0^1\int_0^1 2x^2y^2dxdy + \int_0^1\int_0^1 8x^4y^4dxdy$$

$= \dfrac{2}{9} + \dfrac{8}{25} = \dfrac{50+72}{225} = \dfrac{122}{225}$. Thus Cov$(X, Y) = \dfrac{122}{225} - \left(\dfrac{11}{25}\right)^2 = \dfrac{1}{225}$.

8. Toll rates at a certain bridge is \$ 2 for 2 axis vehicles, \$ 4 for 3 axis vehicles and \$ 10 for vehicles with 4 or more axis. Suppose that 80% of cars passing the bridge have 2 axis, 10% have 3 axis and 10% have 4 or more axis. Let S be the toll collected from next 1 million cars passing the bridge.
(a) Compute ES.
(b) Compute VS.
(c) Compute approximately $P(S > 3003000)$.

Solution. Let X_j be the toll paid by j-th car. Then $S = X_1 + X_2 + \dots X_{1000000}$. Next, $EX = 0.8 * 2 + 0.1 * 2 + 0.1 * 10 = 3$ so $ES = 1000000EX = 3000000$. Likewise $VX = 0.8(2-3)^2 + 0.1(4-3)^2 + 0.1 * (10-3)^2 = 5.8$, so $VS = 1000000 * 5.8 = 5800000$ and $\sigma_S = \sqrt{5800000} =\approx 2408$. By Central Limit Theorem $S \approx N(3000, 2408^2)$, that is $S \approx 3000000 + 2408Z$ where $Z \sim N(0,1)$. Thus

$$P(S > 3003000) \approx P(3000000 + 2408Z > 3003000) = P(2408Z > 3000)$$
$$= P(Z > 1.25) = 1 - P(Z < 1.25) \approx 1 - 0.89 = 0.11.$$

9. Suppose that X and Y are independent random variables with $E(X) = 1, \sigma_X^2 = 2$, $E(Y) = 4, \sigma_Y^2 = 3$. Let $U = 5X - 4Y$.
(a) Find the mean and variance of U.
(b) Compute $Cov(X, U)$.
(c) If X and Y are each normally distributed find $P(U \geq 1)$.

Solution. (a) $EU = 5EX - 4EY = 5 - 16 = -11$. $V(X) = 5^2VX + 4^2VY = 25*2 + 16*3 = 98$.

$$(b) \quad Cov(X,U) = Cov(X, 5X) - Cov(X, 4Y) = 5VX = 10.$$

(c) By part (a) $U \sim N(-11, 98)$. Hence $U = \sqrt{98}Z - 11$ where $Z \sim N(0,1)$. Accoridngly

$$P(U \geq 1) = P(\sqrt{98}Z - 11 \geq 1) = P(\sqrt{98}Z \geq 12) = P\left(Z \geq \frac{12}{\sqrt{98}}\right)$$
$$\approx P(Z \geq 1.21) = 1 - P(Z < 1.21) \approx 0.11.$$

10. Let $X_1, X_2, \dots X_5$ be a sample of some unknown distribution X.
Let $\hat{\mu} = \frac{X_1}{2} + \frac{X_2}{4} + \frac{X_3}{8} + \frac{X_4}{16} + \frac{X_5}{16}$ be an estimator of the population mean.
(a) Is $\hat{\mu}$ unbiased or not?
(b) Express $V(\hat{\mu})$ in terms of $V(X)$.
(c) Compare $\hat{\mu}$ with the sample mean.

Solution.

$$(a) \quad E(\hat{\mu}) = E\left(\frac{X_1}{2}\right) + E\left(\frac{X_2}{4}\right) + E\left(\frac{X_3}{8}\right) + E\left(\frac{X_4}{16}\right) + E\left(\frac{X_5}{16}\right) = EX\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16}\right) = EX$$

so $\hat{\mu}$ is unbiased.

$$(b)\quad V(\hat{\mu}) = V\left(\frac{X_1}{2}\right) + V\left(\frac{X_2}{4}\right) + V\left(\frac{X_3}{8}\right) + V\left(\frac{X_4}{16}\right) + V\left(\frac{X_5}{16}\right)$$
$$= VX\left(\frac{1}{2^2} + \frac{1}{4^2} + \frac{1}{8^2} + \frac{1}{16^2} + \frac{1}{16^2}\right) = 0.3359375VX$$

(c) $V(\bar{X}) = \frac{VX}{5} = 0.2V(X)$. So the sampe mean has a smalller variance and hence is a better estimator.

11. Suppose that W is a discrete random variable such that $P(W=0) = (1-a)/4$, $P(W=1) = (1+2a)/2$, and $P(W=2) = (1-3a)/4$, where $-\frac{1}{2} < a < \frac{1}{3}$ is an unknown parameter.

(a) Compute EW.

Let (2, 0, 1, 1, 1, 2) be the sample from the distribution W.

(b) Find the method of moment estimator for a.

(c) Find the maximum likelyhood estimator of a.

Solution. (a) $EW = 0 * \frac{1-a}{4} + 1 * \frac{1+2a}{2} + 2 * \frac{1-3a}{4} = \frac{2-a}{2}$.

(b) From part (a) we have $\bar{X} = \frac{2-\hat{a}}{2}$ so $\hat{a}_{MOM} = 2 - 2\bar{X} = 2 - \frac{7}{3} = -\frac{1}{3}$.

$$(c)\quad P(2,0,1,1,2,2) = \left(\frac{1-a}{4}\right)\left(\frac{1+2a}{2}\right)^3\left(\frac{1-3a}{4}\right)^2$$

so the likelyhood function is

$$L(a) = \ln(1-a) + 3\ln(1+2a) + 2\ln(1-3a) - 5\ln 4.$$

Thus

$$L'(a) = -\frac{1}{1-a} + \frac{6}{1+2a} - \frac{6}{1-3a}.$$

$L'(a) = 0$ if

$-(1+2a)(1-3a) + 6(1-a)(1-3a) - 6(1-a)(1+2a) = 0$. That is $36a^2 - 29a - 1 = 0$

Roots have form $\dfrac{29 \pm \sqrt{985}}{72}$. Only one of those roots belongs to the parameter interval, so that $\hat{a}_{MLE} = \dfrac{29 - \sqrt{985}}{72} \approx -0.033$.

12. Let X be a random variable with density $f_X(x) = 1 - \theta + 2\theta x$ if $0 \le x \le 1$ (and $f_X(x) = 0$ otherwise) where $0 < \theta < 1$ is an unknown parameter.

(a) Compute EX and VX.

Let $X_1, X_2 \ldots X_n$ be a sample of this distribution.

(b) Find the method of moment estimator for θ.

(c) Give an estimate for the variance of the estimator from part (b).

Solution.

$$(a)\quad EX=\int_0^1(1-\theta+2\theta x)x\,dx=\frac{1-\theta}{2}+\frac{2\theta}{3}=\frac{1}{2}+\frac{\theta}{6}.$$

$$EX^2=\int_0^1(1-\theta+2\theta x)x^2dx=\frac{1-\theta}{3}+\frac{\theta}{2}=\frac{1}{3}+\frac{\theta}{6}.$$

$$VX=\frac{1}{3}+\frac{\theta}{6}-\left(\frac{1}{2}+\frac{\theta}{6}\right)^2=\frac{1}{12}-\frac{\theta^2}{36}.$$

(b) From part (a) we have $\bar{X}=\frac{1}{2}+\frac{\hat{\theta}}{6}$ so that $\hat{\theta}=6\bar{X}-3$.

(c) $V(\hat{\theta})=6^2V(\bar{X})=\frac{36VX}{n}$. So the estimate $\hat{V}(\hat{\theta})=\frac{36}{n}\widehat{VX}$. To estimate the variance of X we can use either $S=\frac{1}{n-1}\sum_{j=1}^n(X_j-\bar{X})^2$ which is always a reasonable estimate for the population variance or a model specific estimate $\frac{1}{12}-\frac{(\hat{\theta})^2}{36}$.

13. After examining 50 bags of rice in a warehouse an inspector came with the following 95% confidence interval for the average weight of a bag 9.9 ± 0.2 lbs.

(a) Compute 99% confidence interval for the average weight of a bag;

(b) Compute 95% lower confidence bound for the average weight of a bag;

(c) Estimate how many bags need to be examined so that the width of the 95% confidence interval is 0.1 lbs.

Solution. 95% confidence interval has form $\bar{X}\pm z_{0.025}\frac{s}{\sqrt{50}}$. Thus $\bar{X}=9.9$ and $\frac{1.96s}{\sqrt{50}}=0.2$. Accordingly $s=\frac{0.2\sqrt{50}}{1.96}=0.72$ Next, 99% confidence interval has form

$$\bar{x}\pm z_{0.995}\frac{s}{\sqrt{n}}=9.9\pm\frac{2.58\times 0.72}{\sqrt{50}}=9.9\pm 0.26=[9.64,10.16].$$

(b) Lower confidence bound is

$$\bar{X}-z_{0.95}\frac{s}{\sqrt{n}}=9.9-\frac{1.65\times 0.72}{\sqrt{50}}=9.9-0.17=9.73.$$

(c) The width of the 95% confidence interval is $2z_{0.025}\frac{s_N}{\sqrt{N}}$. So from equation $\frac{2*1.96s_N}{\sqrt{N}}=0.1$ we get $N=(\frac{2*1.96s_N}{0.1})^2$. Plugging our estimate for $\hat{s}_N=0.72$ we get $N=800$.

14. Researchers investigated the physiological changes that accompany laughter. Ninety (90) subjects ($18-34$ years old) watched film clips desighed to evoke laughter. During the laughing period, the researchers measured the heart rate (beats per minute) of subject, with the following summary results: $\overline{x}=73.1$, $s=3$. It is well known that the mean resting heart rate of adults is 71 beats per minute.

(a) Test whether the true mean heart rate during laughter exceeds 71 beats per minute using $\alpha=.05$.

(b) Find the power of the test at $\mu=72$

(c) Estimate true mean heart rate during laughter in a way that conveys information about precision and reliability. (Calculate 95% CI).

Solution. (a) We have to test $H_0 = \{\mu = 71\}$ vs $H_a = \{\mu > 71\}$. Test statistics is $z = \frac{\bar{X}-71}{s/\sqrt{90}}$. In our case $z = \frac{2.1}{3/\sqrt{90}} = 6.65$. Since $z > z_{0.05} = 1.65$ we have sufficient evidence to reject H_0. That is we have a strong evidence that laughter causes the the average heart rate to exceed 71 beats per minute.

(b) using the formula on page 314 of the book we get

$$\beta = \Phi\left(z_{0.05} + \frac{71-72}{3/\sqrt{90}}\right) = \Phi(-1.51) \approx 0.07.$$

Hence the power is $1 - \beta = 0.93$.

(c) The confidence interval has the form

$$\bar{X} \pm z_{0.05}\frac{s}{\sqrt{n}} = 73.1 \pm \frac{1.96 * 3}{\sqrt{90}} = 73.1 \pm 0.62 = [72.48, 73.72].$$

The upper confidence bound has the form

$$\bar{X} \pm z_{0.025}\frac{s}{\sqrt{n}} = 73.1 \pm \frac{1.96 * 3}{\sqrt{90}} = 73.1 \pm 0.62 = [72.48, 73.72].$$

15. A business journal investigation of the performance and timing of corporate acquisitions discovered that in a random sample of 2,863 firms, 848 announced one or more acquisitions during the year 2000. Let p be the proportion of firms which made one or more acquisitions during the year 2000.

(a) Give a point estimate for p;

(b) Construct 90% confidence interval for p;

(c) Construct 90% upper confidence bound for p.

Solution. (a) $\hat{p} = \frac{X}{n} = \frac{848}{2863} = 0.296$.

(b) The confidence interval takes form

$$\frac{\hat{p} + z_{0.05}^2/(2*2863)}{1 + z_{0.05}^2/(2863)} \pm z_{0.05}\frac{\sqrt{\hat{p}(1-\hat{p})/(2863) + z_{0.05}^2/(4*2863)^2}}{1 + z_{0.05}^2/(2863)} = 0.296 \pm 0.014 = [0.282, 0.310].$$

(b) The upper confidence bound takes form

$$\frac{\hat{p} + z_{0.1}^2/(2*2863)}{1 + z_{0.1}^2/(2863)} + z_{0.1}\frac{\sqrt{\hat{p}(1-\hat{p})/(2863) + z_{0.1}^2/(4*2863)^2}}{1 + z_{0.1}^2/(2863)} = 0.296 \pm 0.011 = 0.307.$$

16. Strategic placement of lobster traps is one of keys for a successful lobster fisherman. A study was conducted of the average distance separating traps by lobster fishermen. The trap-spacing measurements (in meters) for a sample of seven teams from *Blue Sea* fishing cooperative and came with the following data: $\overline{x} = 81.00$ and $s^2 = 125$. Suppose that the trap-spacing has normal distribution.

(a) Test the hypotheis that the average distance between the traps is at least 90m at the significance level $\alpha = 0.05$.

(b) What is the problem with using the normal (z) statistic to find a confidence interval for the average distance between the traps?

(c) Construct 95% confidence interval for the average distance between the traps.

(d) One team from *Blue Sea* cooperative was not working during the day of test. Given 95% prediction interval for the distance between the traps used by the missing team.

Solution. (a) We have to test $H_0 = \{\mu = 90\}$ vs $H_a = \{\mu < 90\}$. The test statistics is $t = \frac{\bar{x}-90}{\sqrt{s^2/7}}$ which in our case equals to -2.130. The ctritiacal value $t_{0.025,6} = 2.447$. Since $|t| < t_{0.025,6}$ we do not reject H_0. That is we do not have sufficient evidence that the average distance between the traps is in fact less that 90 m.

(b) We do not know the standard deviation and since the number of observations is small we can not estimate it accurately enough to use z statistic.

(c) The confidence interval takes form

$$81 \pm 2.447\sqrt{s^2/7} = [70.69, 91.31].$$

Note that 90 is inside the confidence interval which is in agreement with the fact that we did not reject H_0.

(d) The prediction interval takes form

$$81 \pm 2.447\sqrt{s^2}\sqrt{1+\frac{1}{7}} = [51.84, 110.16].$$