Statistics 394, Problem Set 9 Solutions

Wellner; 3/8/2000

1. K, 4.3, # 4, page 269. Let X have the density $f(x) = 2x1_{(0,1)}(x)$.
(a) Find $E(X)$ and $Var(X)$.
(b) Find the kth moment μ_k for $k = 0, 1, 2, \ldots$.

   Solution: This is easiest if we do (b) first:
(b) The kth moment is

$$\begin{aligned} \mu_k &\equiv E(X^k) = \int_0^1 x^k f(x)dx = \int_0^1 x^k(2x)dx = 2\int_0^1 x^{k+1}dx \\ &= \frac{2}{k+2}x^{k+2}|_0^1 = \frac{2}{k+2}\,. \end{aligned}$$

   In particular for $k = 0, 1, 2$, $\mu_0 = 1$, $\mu_1 = 2/3$, $\mu_2 = 2/4 = 1/2$.
(a) By the results from (b) it follows that $E(X) = \mu_1 = 2/3$ and $Var(X) = E(X^2) - (EX)^2 = 1/2 - (2/3)^2 = 1/18$, which agrees with our calculations in HO # 8.

2. K, 4.3, # 14, page 269. If $E(X) = \mu$ and $Var(X) = b$, find the following:
(a) $E(2X - 3)$; (b) $Var(2X - 3)$;
(c) $E(X^2)$; (d) $E(X - 1)^2$;
(e) $Var(5 - X)$; (f) $E[(X - 2)(X + 1)]$.

   Solution: (a) $E(2X - 3) = 2E(X) - 3 = 2\mu - 3$.
(b) $Var(2X - 3) = 4Var(X) = 4b$.
(c) $E(X^2) = Var(X) + (EX)^2 = b + \mu^2$.
(d) $E(X - 1)^2 = E(X^2) - 2E(X) + 1 = b + \mu^2 - 2\mu + 1 = b + (\mu - 1)^2$.
(e) $Var(5 - X) = Var(X) = b$.
(f) $E[(X - 2)(X + 1)] = E(X^2) - E(X) - 2 = b + \mu^2 - \mu - 2$.

3. K, 4.3, # 18, page 269. Find an expression for the variance of XY if X and Y are independent.

   Solution: By our computational formula for the variance,

$$\begin{aligned} Var(XY) &= E[(XY)^2] - [E(XY)]^2 = E[X^2Y^2] - [E(XY)]^2 \\ &= E(X^2)\cdot E(Y^2) - [E(X)\cdot E(Y)]^2 \qquad \text{by independence (twice!)} \end{aligned}$$

   Note that this is *not* equal to the product of the Variances:

$$\begin{aligned} Var(X)Var(Y) &= \{E(X^2) - (EX)^2\}\{E(Y^2) - (EY)^2\} \\ &= E(X^2)\cdot E(Y^2) - (EX)^2E(Y^2) - E(X^2)(EY)^2 + [E(X)\cdot E(Y)]^2\,. \end{aligned}$$

4. K, Supplementary Exercises, Chapter 4, # 8, page 299 Let X be the smaller and Y the larger of the two numbers that show when a pair of fair dice is rolled. (If the two numbers that show are equal, then $X = Y$.)
(a) Make a table of the joint probability mass function of X and Y. Include the marginal mass functions.
(b) Find $E(X)$, $E(X^2)$, EY, $E(Y^2)$, and $E(XY)$.
(c) Find $Var(X)$ and $Var(Y)$.
(d) Find the covariance of X and Y. [Guess first whether it is positive, negative, or 0.]
(e) Find the correlation coefficient of X and Y.

Solution: (a) If we record the values of $X = \min\{Y_1, Y_2\}$ and $Y = \max\{Y_1, Y_2\}$ where Y_1 and Y_2 are the two numbers which appear on the two dice, then we get following table:

y_2/y_1	1	2	3	4	5	6
6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(5,6)
4	(1,4)	(2,4)	(3,4)	(4,4)	(4,5)	(4,5)
3	(1,3)	(2,3)	(3,3)	(3,4)	(3,5)	(3,6)
2	(1,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)

This yields the following joint and marginal distributions for (X, Y): the entries in the table are $36 \cdot p_{X,Y}(x, y)$, $36 \cdot p_X(x)$, and $36 \cdot p_Y(y)$.

Y/X	1	2	3	4	5	6	$p_Y(y)$
6	2	2	2	2	2	1	11
5	2	2	2	2	1		9
4	2	2	2	1			7
3	2	2	1				5
2	2	1					3
1	1						1
$p_X(x)$	11	9	7	5	3	1	36

(b) From the table we can easily compute

$$\begin{aligned} E(X) &= \frac{1}{36}(1 \cdot 11 + 2 \cdot 9 + 3 \cdot 7 + 4 \cdot 5 + 5 \cdot 3 + 6 \cdot 1) \\ &= \frac{1}{36}(11 + 18 + 21 + 20 + 15 + 6) \\ &= \frac{91}{36} = 2.528\,. \end{aligned}$$

Similarly,

$$E(X^2) = \frac{1}{36}(1^2 \cdot 11 + 2^2 \cdot 9 + 3^2 \cdot 7 + 4^2 \cdot 5 + 5^2 \cdot 3 + 6^2 \cdot 1)$$

$$
\begin{aligned}
&= \frac{1}{36}(11+36+63+80+75+36) \\
&= \frac{301}{36} = 8.361\,.
\end{aligned}
$$

To compute $E(Y)$ and $E(Y^2)$, first note that $P(Y=k) = P(X = 7-k)$, $k = 1, \ldots, 6$. Thus Y has the same distribution as $7 - X$. Thus $E(Y) = E(7-X) = 7 - E(X) = 7 - 91/36 = 161/36 = 4.472$, while

$$E(Y^2) = E(7-X)^2 = 49-14E(X)+E(X^2) = 49-14(91)/36+(301)/36 = 791/36 = 21.972\,.$$

Furthermore

$$
\begin{aligned}
E(XY) &= \frac{1}{36}(1\cdot 1 + 1\cdot 2\cdot 2 + 1\cdot 3\cdot 2 + 1\cdot 4\cdot 2 + 1\cdot 5\cdot 2 + 1\cdot 6\cdot 2 + 2\cdot 2 + 2\cdot 3\cdot 2 \\
&\quad + 2\cdot 4\cdot 2 + 2\cdot 5\cdot 2 + 2\cdot 6\cdot 2 \\
&\quad + 3\cdot 3 + 3\cdot 4\cdot 2 + 3\cdot 5\cdot 2 + 3\cdot 6\cdot 2 + 4\cdot 4 + 4\cdot 5\cdot 2 + 4\cdot 6\cdot 2 \\
&\quad + 5\cdot 5 + 5\cdot 6\cdot 2 + 6\cdot 6) \\
&= \frac{1}{36}(1+4+6+8+10+12+4+12+16+20+24 \\
&\quad + 9+24+30+36+16+40+48+25+60+36) \\
&= \frac{441}{36}\,.
\end{aligned}
$$

(c) From the results of (b) we compute

$$Var(X) = E(X^2) - (EX)^2 = \frac{301}{36} - \left(\frac{91}{36}\right)^2 = 1.97145\,.$$

Since Y has the same distribution as $7 - X$ it follows immediately that $Var(Y) = Var(X) = 1.97145$.
(d) From looking at the region where the joint mass function of X, Y is positive, it is natural to guess that $Cov(X,Y) > 0$. In fact,

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{441}{36} - \frac{91}{36}\frac{161}{36} = 0.9452\,.$$

(e) The correlation is

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{.9452}{1.97145} = 0.479\,.$$

5. K, Supplementary Exercises, Chapter 4, # 9, page 299 Suppose $E(X) = 4$, $Var(X) = 3$, $E(Y) = 2$, $Var(Y) = 5$, and $\rho(X,Y) = .32$. Find the following:
(a) $E(X^2)$; (b) $Cov(X,Y)$;
(c) $E(XY)$; (d) $Cov(2X+3, 1-Y)$;
(e) $Var(X+Y)$; (f) $Var(3X-Y)$;
(g) $Cov(X, X+Y)$.

Solution: (a) $E(X^2) = Var(X) + (EX)^2 = 3 + 4^2 = 19$.
(b) $Cov(X,Y) = \rho(X,Y)\sqrt{Var(X)Var(Y)} = .32\sqrt{3\cdot 5} = .32\sqrt{15} = 1.239355$.

(c) $E(XY) = Cov(X,Y) + E(X) \cdot E(Y) = 1.23935 + 4 \cdot 2 = 9.23935$.
(d) $Cov(2X+3, 1-Y) = -2Cov(X,Y) = -2 \cdot 1.23935 = -2.47871$.
(e) $Var(X+Y) = Var(X) + 2Cov(X,Y) + Var(Y) = 3 + 2(1.23935) + 5 = 10.47871$.
(f) $Var(3X - Y) = 9Var(X) - 6Cov(X,Y) + Var(Y) = 9 \cdot 3 - 6(1.23935) + 5 = 24.5639$.
(g) $Cov(X, X+Y) = Cov(X,X) + Cov(X,Y) = Var(X) + Cov(X,Y) = 3 + 1.23935 = 4.23935$.

6. K, 6.2, # 4, page 359. Let $X_1, \ldots, X_{40}$ be a sample of size 40 from the $N(50, 36)$ distribution. Find the following:
(a) The probability that $\overline{X}_{40}$ is between 50 and 51.
(b) The probability that S_{40} is less than 1940.
(c) The 98th percentile of the distribution of $\overline{X}_{40}$.
(d) Two numbers, equidistant from the expected value of S_{40}, so that the probability is .75 that S_{40} is between them.

Solution: (a) Since $\overline{X}_{40} \sim N(50, 36/40)$, it follows that $Z \equiv (\overline{X}_{40} - 50)/\sqrt{36/40} \sim N(0,1)$, and hence

$$\begin{aligned} P(50 < \overline{X}_{40} < 51) &= P\left(0 < \frac{\overline{X}_{40} - 50}{\sqrt{36/40}} < \frac{51 - 50}{\sqrt{36/40}}\right) \\ &= P(0 < Z < 1.0541) = .3531\,. \end{aligned}$$

(b) Now $(S_{40} - 40 \cdot 50)/\sqrt{40 \cdot 36} \sim N(0,1)$, so

$$\begin{aligned} P(S_{40} < 1940) &= P\left(\frac{S_{40} - 2000}{\sqrt{40 \cdot 36}} < \frac{1940 - 2000}{\sqrt{40 \cdot 36}}\right) \\ &= P(Z < -1.581) = P(Z > 1.581) = 1 - P(Z \leq 1.581) \\ &= 1 - .9429 = .0571\,. \end{aligned}$$

(c) Now by definition the 98th percentile $x_{.98}$ of $\overline{X}_{40}$ satisfies

$$P(\overline{X}_{40} < x_{.98}) = .98\,.$$

But since $Z \equiv (\overline{X}_{40} - 50)/\sqrt{36/40} \sim N(0,1)$, we have

$$P(\overline{X}_{40} < x_{.98}) = P\left(\frac{\overline{X}_{40} - 50}{\sqrt{36/40}} < \frac{x_{.98} - 50}{\sqrt{36/40}}\right) = P\left(Z < \frac{x_{.98} - 50}{\sqrt{36/40}}\right)$$

which equals .98 if

$$\frac{x_{.98} - 50}{\sqrt{36/40}} = z_{.98} = 2.055\,.$$

This implies that

$$x_{.98} = 50 + 2.055 \cdot \sqrt{36/40} = 51.94954\,.$$

(d) Since $E(S_{40}) = 40 \cdot 50 = 2000$, we want to find two numbers of the form $2000 - a$ and $2000 + a$ so that

$$P(2000 - a < S_{40} < 2000 + a) = .75\,.$$

But since $(S_{40} - 40 \cdot 50)/\sqrt{40 \cdot 36} \sim N(0,1)$, the probability on the left side is

$$= P\left(\frac{-a}{\sqrt{40 \cdot 36}} < Z < \frac{a}{\sqrt{40 \cdot 36}}\right) = 2P\left(0 < Z < \frac{a}{\sqrt{40 \cdot 36}}\right) = .75$$

if

$$\frac{a}{\sqrt{40 \cdot 36}} = z_{.875} \doteq 1.15\,,$$

and this implies that $a = 1.15 \times \sqrt{40 \cdot 36} = 43.63943$.

7. K, 6.2, # 5, page 359. A company sells sugar in 5-pound bags; it is subject to a penalty for each bag it sells that is found to contain less than 5 pounds of sugar. The packaging machines cannot guarantee an exact weight of sugar per bag; in fact, the weights are normally distributed with an average weight of $\mu = 5.13$ pounds and a standard deviation of $\sigma = .08$ pound.
(a) What proportion of bags weigh less than 5 pounds?
(b) The machine can be adjusted to increase the average weight μ. Assume that the standard deviation will not change. To what value should μ be set so that only 1% of the bags will weigh less than 5 pounds?

Solution: (a) Note that $Z \equiv (X - \mu)/\sigma \sim \mathrm{N}(0,1)$. Thus

$$\begin{aligned} P(X < 5) &= P((X-\mu)/\sigma < (5-\mu)/\sigma) = P(Z < (5 - 5.13)/.08) \\ &= P(Z < -1.625) = P(Z > 1.625) = 1 - P(Z \leq 1.625) = 1 - .9479 = .0521\,. \end{aligned}$$

(My answer differs slightly from Kelly's answer of .0516 because I interpolated between the standard normal distribution at the points 1.62 and 1.63 in the table on page 607; Kelly apparently rounded up to 1.63.)
(b) To find the value of μ so that $P(X < 5) = .01$, we first note that $P(Z < -2.33) = .0099 \doteq .01$ from the table on page 607. Thus it follows that

$$\begin{aligned} .01 \doteq P(Z < -2.33) &= P((X-\mu)/\sigma < -2.33) \\ &= P(X < \mu - 2.33(.08)) = P(X < 5) \end{aligned}$$

if $\mu - 2.33(.08) = 5$, or $\mu = 5 + (2.33)(.08) = 5.1864$.

8. K, 6.2, # 6, page 359. Heights of children in a certain age group average 58.4 inches, with a standard deviation of 2.9 inches. Assume that the heights are normally distributed.
(a) What proportion of children are between 57 and 61 inches tall?
(b) What is the 90th percentile of the children's heights?
(c) What is the probability that the average of a sample of 20 heights will be greater than 60 inches?

Solution: (a) Since $(X - \mu)/\sigma \sim N(0,1)$,

$$\begin{aligned} P(57 < X < 61) &= P((57 - 58.4)/2.9 < Z < (61 - 58.4)/2.9) = P(-.483 < Z < .8965) \\ &= \Phi(.8965) - \Phi(-.483) \\ &= .8149 - 0.314 = 0.5009\,. \end{aligned}$$

(b) We want to find a number, $x_{.90}$ which satisfies $P(X < x_{.90}) = .90$. But since $(X - \mu)/\sigma \sim N(0, 1)$,

$$P(X < x_{.90}) = P(Z < (x_{.90} - 58.4)/2.9) = .90$$

if

$$(x_{.90} - 58.4)/2.9 = z_{.90} \doteq 1.18\,.$$

Thus $x_{.90} = 58.4 + (2.9)(1.28) = 62.112$.
(c) Since $\overline{X}_20 \sim N(58.4, (2.9)^2/20)$,

$$\begin{aligned} P\left(\overline{X}_{20} > 60\right) &= P\left(\frac{\overline{X}_{20} - 58.4}{2.9/\sqrt{20}} > \frac{60 - 58.4}{2.9\sqrt{20}}\right) \\ &= P(Z > 2.467) = .0068\,. \end{aligned}$$

9. K, 6.3, # 4, page 370; plus do part (c) another way. Let S_{45} be the number of failures preceding the 45th success in a Bernoulli process with success probability $p = .36$. We can write $S_{45} = X_1 + \cdots + X_{45}$, where X_1 is the number of failures preceding the first success, X_2 is the number of failures between the first and second successes, and so forth. The X_j's are independent.
(a) Give the name of the distribution of a single X_j; also give its expected value μ and its variance σ^2.
(b) What are the expected value and variance of S_{45}?
(c) Approximate the probability that S_{45} is within 20 of its expected value.
(d) What does Chebychev's inequality say about the probability you approximated in part c?

Solution: (a) The X_j's have a geometric distribution with parameter p: $P(X_j = k) = q^k p$ for $k = 0, 1, 2, \ldots$, $j = 1, 2, \ldots$. Recall that the geometric distribution of Handout #4 on the Bernoulli process is the distribution of the *number of trials* Y_j until the next success, and hence $Y_j = X_j + 1$. Thus we have $P(Y_j = k) = P(X_j = k - 1) = q^{k-1}p$ for $k = 1, 2, \ldots$.
Now as we calculated in Problem 8.8 (Bonus problem 1, K 4.2, # 9, page 257), $E(X_1) = E(Y_1) - 1 = (1/p) - 1 = q/p$, and $Var(X_1) = Var(Y_1) = q/p^2$.
(b) $S_{45} = \sum_{j=1}^{45} X_j$, so $E(S_{45}) = 45 \cdot q/p = 45(.64)/.36 = 80$, $Var(S_{45}) = 45q/p^2 = 45(.64)/(.36)^2 = 222.22$.
(c) By the CLT (with correction for continuity),

$$\begin{aligned} P(80 - 20 \le S_{45} \le 80 + 20) &= P(60 - .5 \le S_{45} \le 100 + .5) = P(59.5 \le S_{45} \le 100.5) \\ &= P\left(\frac{59.5 - 80}{\sqrt{222.22}} < \frac{S_{45} - 80}{\sqrt{222.22}} < \frac{100.5 - 80}{\sqrt{222.22}}\right) \\ &\doteq P(-1.375 \le Z \le 1.375) = 2(.4155) = .8310\,. \end{aligned}$$

If I do the same calculation without the correction for continuity I get $P(-1.34 \le Z \le 1.34) = 2(.4099) = 0.8198$ which agrees with Kelly's answer. Now we'll do this problem another way: first note that $S_{45} = \sum_{j=1}^{45} X_j = \sum_{j=1}^{45}(Y_j - 1) =$

$\sum_{j=1}^{45} Y_j - 45 \equiv W_{45} - 45$ where we know, using the basic identity for the Bernoulli process, that $[W_{45} > k] = [T_k < 45]$. Thus it follows that

$$\begin{aligned}
P(60 \le S_{45} \le 100) &= P(59 < W_{45} - 45 < 101) = P(104 < W_{45} < 146) \\
&= P(W_{45} > 104) - P(W_{45} > 145) \\
&= P(T_{104} < 45) - P(T_{145} < 45) \qquad \text{where } T_k \sim \text{Binomial}(k, p) \\
&= P(T_{104} \le 44.5) - P(T_{145} \le 44.5) \qquad \text{(continuity correction)} \\
&= P\left(\frac{T_{104} - 104(.36)}{\sqrt{104(.36)(.64)}} < \frac{44.5 - 104(.36)}{\sqrt{104(.36)(.64)}}\right) \\
&\qquad - P\left(\frac{T_{145} - 145(.36)}{\sqrt{145(.36)(.64)}} < \frac{44.5 - 145(.36)}{\sqrt{145(.36)(.64)}}\right) \\
&\doteq P(Z < 1.442) - P(Z < -1.332) = .9254 - (1 - .9085) = .834\,.
\end{aligned}$$

(d) Chebychev's inequality says

$$P(80 - 20 < S_{45} < 80 + 20) \ge 1 - \frac{Var(S_{45})}{20^2} = 1 - \frac{45(.64)/(.36)^2}{400} = 1 - .556 = .444\,.$$

10. Bonus Problem 1 (and Practice Problem for the Final Exam (Highly Recommended): K, 6.4, # 7 and # 8, page 382. # 7. Use the normal approximation to the binomial distribution to find the approximate probability that in 79 tosses of a fair coin there are fewer than 30 heads. Use a continuity correction.

    No solution here; this one is on the practice final. It is closely related to the "second" or "other" solution to part (c) of the previous problem.

11. Bonus Problem 2. Under "Statistics; Random Samples" at the virtual laboratories web site, **http://www.math.uah.edu/stat/sample/index.html**, go to the **sample mean experiment**.
    A. Look at the experiment for the normal distribution with $\mu = 0$, $d = \sigma = 1$, and $n = 4$. Describe what you see and explain it. Change μ to 5 and n to 16. Describe what you see (the blue curves) and explain it.
    B. Run the normal experiment with stop frequency 100, $\mu = 0$, $d = \sigma = 1$, and $n = 4$. Describe what you see (the red curves) and explain it.
    C. Look at the binomial part of the experiment with stop frequency 100, $m = 2$, $p = .75$, and $n = 4$. Explain what you see (the blue bars).
    D. Run the binomial part of the experiment with stop frequency 100, $m = 2$, $p = .75$, and $n = 4$ and explain what you see (the red bars).

    Solution: A solution to this problem will be posted later.