Stat 110 Strategic Practice 8 Solutions, Fall 2011

1 Covariance and Correlation

1. Two fair six-sided dice are rolled (one green and one orange), with outcomes X and Y respectively for the green and the orange.

   (a) Compute the covariance of $X + Y$ and $X - Y$.

   $\text{Cov}(X + Y, X - Y) = \text{Cov}(X, X) - \text{Cov}(X, Y) + \text{Cov}(Y, X) - \text{Cov}(Y, Y) = 0.$

   (b) Are $X + Y$ and $X - Y$ independent? Show that they are, or that they aren't (whichever is true).

   They are not independent: information about $X + Y$ may give information about $X - Y$, as shown by considering an *extreme example*. Note that if $X + Y = 12$, then $X = Y = 6$, so $X - Y = 0$. Therefore, $P(X - Y = 0|X + Y = 12) = 1 \neq P(X - Y = 0)$, which shows that $X + Y$ and $X - Y$ are not independent. Alternatively, note that $X + Y$ and $X - Y$ are both even or both odd, since the difference $X + Y - (X - Y) = 2Y$ is even.

2. A chicken lays a Poisson(λ) number N of eggs. Each egg, independently, hatches a chick with probability p. Let X be the number which hatch, so $X|N \sim \text{Bin}(N, p)$.

   Find the correlation between N (the number of eggs) and X (the number of eggs which hatch). Simplify; your final answer should work out to a simple function of p (the λ should cancel out).

   As shown in class, in this story X is independent of Y, with $X \sim \text{Pois}(\lambda p)$ and $Y \sim \text{Pois}(\lambda q)$, for $q = 1 - p$. So

   $$\text{Cov}(N, X) = \text{Cov}(X + Y, X) = \text{Cov}(X, X) + \text{Cov}(Y, X) = \text{Var}(X) = \lambda p,$$

   giving

   $$\text{Corr}(N, X) = \frac{\lambda p}{SD(N)SD(X)} = \frac{\lambda p}{\sqrt{\lambda \lambda p}} = \sqrt{p}.$$

3. Let X and Y be standardized r.v.s (i.e., marginally they each have mean 0 and variance 1) with correlation $\rho \in (-1, 1)$. Find a, b, c, d (in terms of ρ) such that $Z = aX + bY$ and $W = cX + dY$ are uncorrelated but still standardized.

   Let us look for a solution with $Z = X$, finding c and d to make Z and W uncorrelated. By bilinearity of covariance,

   $$\mathrm{Cov}(Z, W) = \mathrm{Cov}(X, cX + dY) = \mathrm{Cov}(X, cX) + \mathrm{Cov}(X, dY) = c + d\rho = 0.$$

   Also, $\mathrm{Var}(W) = c^2 + d^2 + 2cd\rho = 1$. Solving for c, d gives

   $$a = 1, b = 0, c = -\rho/\sqrt{1-\rho^2}, d = 1/\sqrt{1-\rho^2}.$$

4. Let $(X_1, \dots, X_k)$ be Multinomial with parameters n and $(p_1, \dots, p_k)$. Use indicator r.v.s to show that $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$.

   First let us find $\mathrm{Cov}(X_1, X_2)$. Consider the story of the Multinomial, where n objects are being placed into categories $1, \dots, k$. Let I_i be the indicator r.v. for object i being in category 1, and let J_j be the indicator r.v. for object j being in category 2. Then $X_1 = \sum_{i=1}^n I_i, X_2 = \sum_{j=1}^n J_j$. So

   $$\begin{aligned} \mathrm{Cov}(X_1, X_2) &= \mathrm{Cov}(\sum_{i=1}^n I_i, \sum_{j=1}^n J_j) \\ &= \sum_{i,j} \mathrm{Cov}(I_i, J_j). \end{aligned}$$

   All the terms here with $i \neq j$ are 0 since the ith object is categorized independently of the jth object. So this becomes

   $$\sum_{i=1}^n \mathrm{Cov}(I_i, J_i) = n\mathrm{Cov}(I_1, J_1) = -np_1p_2,$$

   since

   $$\mathrm{Cov}(I_1, J_1) = E(I_1J_1) - (EI_1)(EJ_1) = -p_1p_2.$$

   By the same method, we have $\mathrm{Cov}(X_i, X_j) = -np_ip_j$ for all $i \neq j$.

5. Let X and Y be r.v.s. Is it correct to say "$\max(X, Y) + \min(X, Y) = X + Y$? Is it correct to say "$\mathrm{Cov}(\max(X, Y), \min(X, Y)) = \mathrm{Cov}(X, Y)$ since either the max is X and the min is Y or vice versa, and covariance is symmetric"?

The identity $\max(x, y) + \min(x, y) = x + y$ is true for all numbers x and y. The random variable $M = \max(X, Y)$ is *defined* by $M(s) = \max(X(s), Y(s))$; this just says to perform the random experiment, observe the numerical values of X and Y, and take their maximum. It follows that

$$\max(X, Y) + \min(X, Y) = X + Y$$

for all r.v.s X and Y, since whatever the outcome s of the random experiment is, we have

$$\max(X(s), Y(s)) + \min(X(s), Y(s)) = X(s) + Y(s).$$

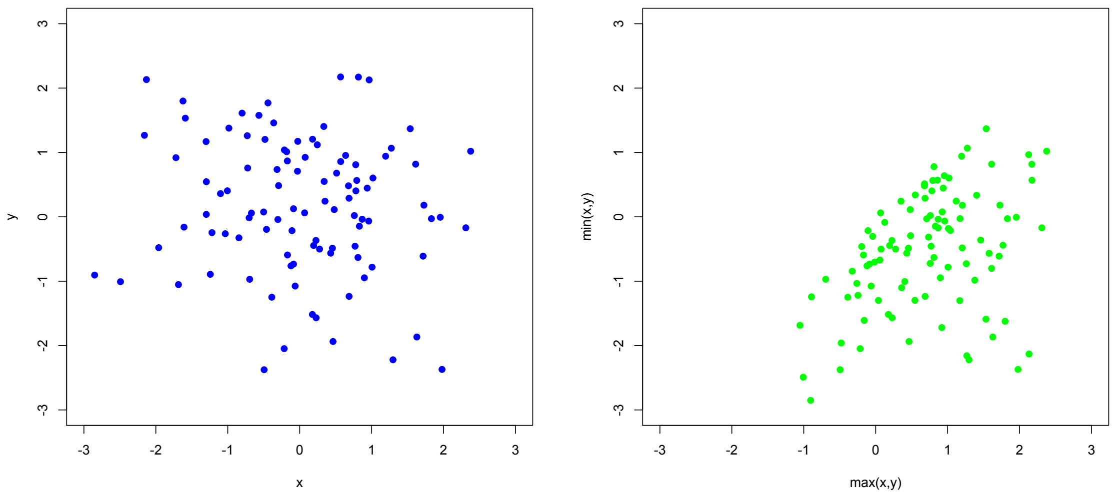
In contrast, the covariance of two r.v.s is a number, not a r.v.; it is *not* defined by observing the values of the two r.v.s and then taking their covariance (that would be a useless quantity, since the covariance between two numbers is 0). It is wrong to say "$\mathrm{Cov}(\max(X, Y), \min(X, Y)) = \mathrm{Cov}(X, Y)$ since either the max is X and the min is Y or vice versa, and covariance is symmetric" since the r.v. X does not equal the r.v. $\max(X, Y)$, nor does it equal the r.v. $\min(X, Y)$.

To gain more intuition into this, consider a "repeated sampling interpretation," where we independently repeat the same experiment many times and observe pairs $(x_1, y_1), \ldots, (x_n, y_n)$, where (x_j, y_j) is the observed value of (X, Y) for the $j$the experiment. Suppose that X and Y are independent non-constant r.v.s (and thus they are uncorrelated). Imagine a *scatter plot* of the observations (which is just a plot of the points (x_j, y_j)). Since X and Y are independent, there should be no pattern or trend in the plot.

On the other hand, imagine a scatter plot of the $(\max(x_j, y_j), \min(x_j, y_j))$ points. Here we'd expect to see a clear increasing trend (since the max is always bigger than or equal to the min, so having a large value of the min (relative to its mean) should make it more likely that we'll have a large value of the max (relative to its mean). So it makes sense that $\max(X, Y)$ and $\min(X, Y)$ should be positive correlated. This is illustrated in the plots below, in which we generated $(X_1, Y_1), \ldots, (X_{100}, Y_{100})$ with the X_i's and Y_j's i.i.d. $\mathcal{N}(0, 1)$.

The simulation was done in R, which is free, extremely powerful statistics software available at http://www.r-project.org/, using the following code:

```
x <- rnorm(100); y <- rnorm(100)
plot(x,y, xlim=c(-3,3),ylim=c(-3,3),  col="blue", pch=19)
plot(pmax(x,y),pmin(x,y), xlim=c(-3,3),ylim=c(-3,3), xlab="max(x,y)",
ylab = "min(x,y)", col="green", pch=19)
```



6. Consider the following method for creating a *bivariate Poisson* (a joint distribution for two r.v.s such that both marginals are Poissons). Let $X = V + W, Y = V + Z$ where V, W, Z are i.i.d. Pois(λ) (the idea is to have something borrowed and something new but not something old or something blue).

   (a) Find Cov(X, Y).

   By bilinearity of covariance,

   $$\text{Cov}(X,Y) = \text{Cov}(V,V) + \text{Cov}(V,Z) + \text{Cov}(W,V) + \text{Cov}(W,Z) = \text{Var}(V) = \lambda.$$

   (b) Are X and Y independent? Are they conditionally independent given V?

   Since X and Y are correlated (with covariance $\lambda > 0$), they are not independent. Alternatively, note that $E(Y) = 2\lambda$ but $E(Y|X = 0) = \lambda$ since if $X = 0$ occurs then $V = 0$ occurs. But X and Y are conditionally independent given V, since the conditional joint PMF is

   $$\begin{aligned} P(X = x, Y = y|V = v) &= P(W = x - v, Z = y - v|V = v) \\ &= P(W = x - v, Z = y - v) \\ &= P(W = x - v)P(Z = y - v) \\ &= P(X = x|V = v)P(Y = y|V = v). \end{aligned}$$

This makes sense intuitively since if we observe that $V = v$, then X and Y are the independent r.v.s W and Z, shifted by the constant v.

(c) Find the joint PMF of X, Y (as a sum).

By (b), a good strategy is to condition on V:

$$\begin{aligned}
P(X = x, Y = y) &= \sum_{v=0}^{\infty} P(X = x, Y = y|V = v)P(V = v) \\
&= \sum_{v=0}^{\min(x,y)} P(X = x|V = v)P(Y = y|V = v)P(V = v) \\
&= \sum_{v=0}^{\min(x,y)} e^{-\lambda}\frac{\lambda^{x-v}}{(x-v)!}e^{-\lambda}\frac{\lambda^{y-v}}{(y-v)!}e^{-\lambda}\frac{\lambda^{v}}{v!} \\
&= e^{-3\lambda}\lambda^{x+y}\sum_{v=0}^{\min(x,y)} \frac{\lambda^{-v}}{(x-v)!(y-v)!v!},
\end{aligned}$$

for x and y nonnegative integers. Note that we sum only up to $\min(x, y)$ since we know for sure that $V \leq X$ and $V \leq Y$.

Miracle check: note that $P(X = 0, Y = 0) = P(V = 0, W = 0, Z = 0) = e^{-3\lambda}$.

7. Let X be Hypergeometric with parameters b, w, n.

(a) Find $E\binom{X}{2}$ *by thinking*, without any complicated calculations.

In the story of the Hypergeometric, $\binom{X}{2}$ is the number of pairs of draws such that both balls are white. Creating an indicator r.v. for each pair, we have

$$E\binom{X}{2} = \binom{n}{2}\frac{w}{w+b}\frac{w-1}{w+b-1}.$$

(b) Use (a) to get the variance of X, confirming the result from class that

$$\mathrm{Var}(X) = \frac{N-n}{N-1}npq,$$

where $N = w + b, p = w/N, q = 1 - p$.

By (a),

$$EX^2 - EX = E(X(X-1)) = n(n-1)p\frac{w-1}{N-1},$$

so

$$\begin{aligned}
\text{Var}(X) &= E(X^2) - (EX)^2 \\
&= n(n-1)p\frac{w-1}{N-1} + np - n^2p^2 \\
&= np\left(\frac{(n-1)(w-1)}{N-1} + 1 - np\right) \\
&= np\left(\frac{nw - w - n + N}{N-1} - \frac{nw}{N}\right) \\
&= np\left(\frac{Nnw - Nw - Nn + N^2 - Nnw + nw}{N(N-1)}\right) \\
&= np\left(\frac{(N-n)(N-w)}{N(N-1)}\right) \\
&= \frac{N-n}{N-1}npq.
\end{aligned}$$

2 Transformations

1. Let $X \sim \text{Unif}(0,1)$. Find the PDFs of X^2 and $\sqrt{X}$.

   (PDF of X^2.) Let $Y = X^2$, $0 < x < 1$, and $y = x^2$, so $x = \sqrt{y}$. The absolute Jacobian is $\left|\frac{dx}{dy}\right| = \left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{2\sqrt{y}}$ for $0 < y < 1$. The PDF of Y for $0 < y < 1$ is

   $$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{2\sqrt{y}}$$

   with $f_Y(y) = 0$ otherwise. This is the Beta$(\frac{1}{2}, 1)$ PDF, so $Y = X^2 \sim \text{Beta}(\frac{1}{2}, 1)$.

   (PDF of $\sqrt{X}$.) Now let $Y = X^{1/2}$, $0 < x < 1$, and $y = x^{1/2}$, so $x = y^2$. The absolute Jacobian is $\left|\frac{dx}{dy}\right| = |2y| = 2y$ for $0 < y < 1$. The PDF of Y for $0 < y < 1$ is

   $$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = 2y$$

with $f_Y(y) = 0$ otherwise. This says that Y has a Beta(2,1) distribution.

In general, the same method shows that if X has a Unif(0,1) distribution and $\alpha > 0$, then $X^{\frac{1}{\alpha}}$ has a Beta(α,1) distribution.

2. Let $U \sim \text{Unif}(0, 2\pi)$ and let $T \sim \text{Expo}(1)$ be independent of U. Define $X = \sqrt{2T}\cos U$ and $Y = \sqrt{2T}\sin U$. Find the joint PDF of (X, Y). Are they independent? What are their marginal distributions?

   The joint PDF of U and T is

$$f_{U,T}(u,t) = \frac{1}{2\pi}e^{-t},$$

   for $u \in (0, 2\pi)$ and $t > 0$. Thinking of (X, Y) as a point in the (x, y)-plane, $X^2 + Y^2 = 2T(\cos^2(U) + \sin^2(U)) = 2T$ is the squared distance from the origin and U is the angle. To make the change of variables, we need the Jacobian:

$$\begin{aligned} J = \left|\frac{\partial(x,y)}{\partial(u,t)}\right| &= \begin{vmatrix} -\sqrt{2t}\sin(u) & (2t)^{-1/2}\cos(u) \\ \sqrt{2t}\cos(u) & (2t)^{-1/2}\sin(u) \end{vmatrix} \\ &= -\sin^2(u) - \cos^2(u) \\ &= -1. \end{aligned}$$

   Then

$$\begin{aligned} f_{X,Y}(x,y) &= f_{U,T}(u,t)|J|^{-1} \\ &= \frac{1}{2\pi}\exp\left(\frac{-(x^2+y^2)}{2}\right)|-1|^{-1} \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-y^2}{2}\right). \end{aligned}$$

   This factors into a function of x times a function of y, so X and Y are independent, and they each have the $\mathcal{N}(0, 1)$ distribution. Thus, X and Y are i.i.d. standard Normal r.v.s; this result is called the *Box-Muller* method for generating Normal r.v.s.

3. Let X and Y be independent, continuous r.v.s with PDFs f_X and f_Y respectively, and let $T = X + Y$. Find the joint PDF of T and X, and use this to give an alternative proof that $f_T(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)dx$, a result obtained in class using the law of total probability.

Consider the transformation from (x, y) to (t, w) given by $t = x+y$ and $w = x$. (It may seem redundant to make up the new name "w" for x, but this makes it easier to distinguish between the "old" variables x, y and the "new" variables t, w.) Correspondingly, consider the transformation from (X, Y) to (T, W) given by $T = X + Y, W = X$. The Jacobian matrix is

$$\frac{\partial(t, w)}{\partial(x, y)} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

which has absolute determinant equal to 1. Thus, the joint PDF of T, W is

$$f_{T,W}(t, w) = f_{X,Y}(x, y) = f_X(x)f_Y(y) = f_X(w)f_Y(t-w),$$

and the marginal PDF of T is

$$f_T(t) = \int_{-\infty}^{\infty} f_{T,W}(t, w)dw = \int_{-\infty}^{\infty} f_X(x)f_Y(t-x)dx.$$

3 Existence

1. Let S be a set of binary strings $a_1 \dots a_n$ of length n (where juxtaposition means concatenation). We call S *k-complete* if for any indices $1 \le i_1 < \dots < i_k \le n$ and any binary string $b_1 \dots b_k$ of length k, there is a string $s_1 \dots s_n$ in S such that $s_{i_1} s_{i_2} \dots s_{i_k} = b_1 b_2 \dots b_k$. For example, for $n = 3$, the set $S = \{001, 010, 011, 100, 101, 110\}$ is 2-complete since all 4 patterns of 0's and 1's of length 2 can be found in any 2 positions. Show that if $\binom{n}{k}2^k(1-2^{-k})^m < 1$, then there exists a k-complete set of size at most m.

   Generate m random strings of length n independently, using fair coin flips to determine each bit. Let S be the resulting random set of strings. If we can show that the probability that S is k-complete is *positive*, then we know that a k-complete set of size at most m must *exist*. Let A be the event that S is k-complete. Let $N = \binom{n}{k}2^k$ and let $A_1, \dots, A_N$ be the events of the form "S contains a string which is $b_1 \dots b_k$ at coordinates $i_1 < \dots < i_k$," in any fixed order. For example, if $k = 3$ then A_1 could be the event "S has an element which is 110 at positions $1, 2, 3$." Then $P(A) > 0$ since

$$P(A^c) = P(\cup_{j=1}^N A_j^c) \le \sum_{j=1}^N P(A_j^c) = N(1-2^{-k})^m < 1.$$

2. A hundred students have taken an exam consisting of 8 problems, and for each problem at least 65 of the students got the right answer. Show that there exist two students who collectively got everything right, in the sense that for each problem, at least one of the two got it right.

   Say that the "score" of a pair of students is how many problems at least one of them got right. The expected score of a random pair of students (with all pairs equally likely) is at least $8(1-0.35^2) = 7.02$, as seen by creating an indicator r.v. for each problem for the event that at least one student in the pair got it right. (We can also improve the 0.35^2 to $\frac{35}{100} \cdot \frac{34}{99}$ since the students are sampled without replacement.) So some pair of students must have gotten a score of at least 7.02, which means that they got a score of 8! ($\leftarrow$ not a factorial.)

3. The circumference of a circle is colored with red and blue ink such that 2/3 of the circumference is red and 1/3 is blue. Prove that no matter how complicated the coloring scheme is, there is a way to inscribe a square in the circle such that at least three of the four corners of the square touch red ink.

   Consider a random square, obtained by picking a uniformly random point on the circumference and inscribing a square with that point a corner; say that the corners are $U_1, \dots, U_4$, in clockwise order starting with the initial point chosen. Let I_j be the indicator r.v. of U_j touching red ink. By symmetry, $E(I_j) = 2/3$ so by linearity, the expected number of corners touching red ink is 8/3. Thus, there must exist an inscribed square with at least 8/3 of its corners touching red ink. Such a square must have at least 3 of its corners touching red ink.

4. Ten points in the plane are designated. You have ten circular coins (of the same radius). Show that you can position the coins in the plane (without stacking them) so that all ten points are covered.

   Hint: consider a honeycomb tiling as in http://mathworld.wolfram.com/Honeycomb.html. You can use the fact from geometry that if a circle is inscribed in a hexagon then the ratio of the area of the circle to the area of the hexagon is $\frac{\pi}{2\sqrt{3}} > 0.9$.

   Take a uniformly random honeycomb tiling (to do this, start with any honeycomb tiling and then shift it horizontally and vertically by uniformly random amounts; by periodicity there is a bound on how large the shifts need to be). Choose the tiling so that a circle the same size as one of the coins can be inscribed in each hexagon. Then inscribe a circle in each hexagon, and let I_j be the indicator r.v. for the jth point being contained inside one the circles.

We have $E(I_j) > 0.9$ by the geometric fact mentioned above, so by linearity $E(I_1 + \cdots + I_{10}) > 9$. Thus, there is a positioning of the honeycomb tiling such that all 10 points are contained inside the circles. Putting coins on top of the circles containing the points, we can cover all ten points.