Stat 110 Strategic Practice 8, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

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 $X - Y$.

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

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).

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.

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

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 "$\text{Cov}(\max(X, Y), \min(X, Y)) = \text{Cov}(X, Y)$ since either the max is X and the min is Y or vice versa, and covariance is symmetric"?

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 $\text{Cov}(X, Y)$.

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

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

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

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

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

2 Transformations

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

2. Let $U \sim \mathrm{Unif}(0, 2\pi)$ and let $T \sim \mathrm{Expo}(1)$ 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?

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.

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 $s_{i_1}s_{i_2}\dots s_{i_k} = b_1b_2\dots b_k$. 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.

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.

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.

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 $\frac{\pi}{2\sqrt{3}} > 0.9$.

Stat 110 Strategic Practice 8 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

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,

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

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

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

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

   First let us find $\text{Cov}(X_1, X_2)$. Consider the story of the Multinomial, where n objects are being placed into categories $1, \ldots, k$. 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

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

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

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

   since

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

   By the same method, we have $\text{Cov}(X_i, X_j) = -np_ip_j$ 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 "$\text{Cov}(\max(X, Y), \min(X, Y)) = \text{Cov}(X, Y)$ 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

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

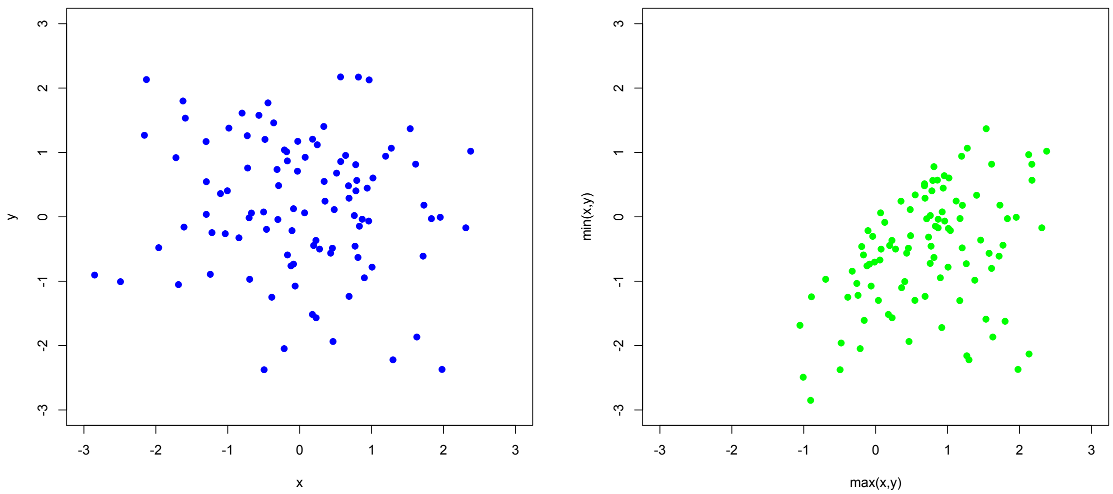
In contrast, the covariance of two r.v.s is a number, not a r.v.; it is *not* defined by observing the values of the two r.v.s and then taking their covariance (that would be a useless quantity, since the covariance between two numbers is 0). It is wrong to say "$\text{Cov}(\max(X, Y), \min(X, Y)) = \text{Cov}(X, Y)$ 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

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

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 $s_{i_1}s_{i_2} \dots s_{i_k} = b_1b_2 \dots b_k$. 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 $U_1, \ldots, U_4$, 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.

Stat 110 Homework 8, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Let X be the number of distinct birthdays in a group of 110 people (i.e., the number of days in a year such that at least one person in the group has that birthday). Under the usual assumptions (no Feb 29, all the other 365 days of the year are equally likely, and the day when one person is born is independent of the days when the other people are born), find the mean and variance of X.

2. A scientist makes two measurements, considered to be independent standard Normal r.v.s. Find the correlation between the larger and smaller of the values.

Hint: note that $\max(x, y) + \min(x, y) = x + y$ and $\max(x, y) - \min(x, y) = |x - y|$.

3. Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \dots$ i.i.d. with a continuous distribution. We say that the jth jumper set a *record* if X_j is greater than all of $X_{j-1}, \dots, X_1$.

(a) Find the variance of the number of records among the first n jumpers (as a sum). What happens to the variance as $n \to \infty$?

(b) A *double record* occurs at time j if *both* the jth and $(j-1)$st jumpers set records. Find the mean number of double records among the first n jumpers (simplify fully; it may help to note that $\frac{1}{j(j-1)} = \frac{1}{j-1} - \frac{1}{j}$). What happens to the mean as $n \to \infty$?

4. Find the PDF of Z^4 for $Z \sim \mathcal{N}(0, 1)$.

5. Let X and Y be independent positive r.v.s, with PDFs f_X and f_Y respectively, and consider the product $T = XY$. When asked to find the PDF of T, Jacobno argues that "it's like a convolution, with a product instead of a sum. To have $T = t$ we need $X = x$ and $Y = t/x$ for some x; that has probability $f_X(x)f_Y(t/x)$, so summing up these possibilities we get that the PDF of T is $\int_0^\infty f_X(x)f_Y(t/x)dx$." Evaluate Jacobno's argument, while getting the PDF of T (as an integral) in 2 ways:

(a) using the continuous law of total probability to get the CDF, and then taking the derivative (you can assume that swapping the derivative and integral is valid);

(b) by taking the log of both sides of $T = XY$ and doing a convolution (and then converting back to get the PDF of T).

6. Let X, Y be continuous r.v.s with a *spherically symmetric* joint distribution, which means that the joint PDF is of the form $f(x, y) = g(x^2 + y^2)$ for some function g. Let

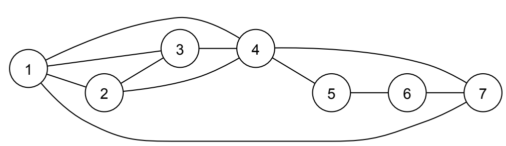
(R, θ) be the polar coordinates of (X, Y), so $R^2 = X^2 + Y^2$ is the squared distance from the origin and θ is the angle (in $[0, 2\pi)$), with $X = R\cos\theta, Y = R\sin\theta$.

(a) Explain intuitively why R and θ are independent. Then prove this by finding the joint PDF of (R, θ).

(b) What is the joint PDF of (R, θ) when (X, Y) is Uniform in the unit disc $\{(x, y) : x^2 + y^2 \le 1\}$?

(c) What is the joint PDF of (R, θ) when X and Y are i.i.d. $\mathcal{N}(0, 1)$?

7. A *network* consists of n *nodes*, each pair of which may or may not have an *edge* joining them. For example, a social network can be modeled as a group of n nodes (representing people), where an edge between i and j means they know each other. Assume the network is undirected and does not have edges from a node to itself (for a social network, this says that if i knows j, then j knows i and that, contrary to Socrates' advice, a person does not know himself or herself). A *clique* of size k is a set of k nodes where every node has an edge to every other node (i.e., within the clique, everyone knows everyone). An *anticlique* of size k is a set of k nodes where there are no edges between them (i.e., within the anticlique, no one knows anyone else). For example, the picture below shows a network with nodes labeled $1, 2, \ldots, 7$, where $\{1, 2, 3, 4\}$ is a clique of size 4, and $\{3, 5, 7\}$ is an anticlique of size 3.



(a) Form a random network with n nodes by independently flipping fair coins to decide for each pair $\{x, y\}$ whether there is an edge joining them. Find the expected number of cliques of size k (in terms of n and k).

(b) A *triangle* is a clique of size 3. For a random network as in (a), find the variance of the number of triangles (in terms of n).

Hint: find the covariances of the indicator random variables for each possible clique. There are $\binom{n}{3}$ such indicator r.v.s, some pairs of which are dependent.

(c) Suppose that $\binom{n}{k} < 2^{\binom{k}{2}-1}$. Show that there is a network with n nodes containing no cliques of size k or anticliques of size k.

Hint: explain why it is enough to show that for a random network with n nodes, the probability of the desired property is positive; then consider the complement.

Stat 110 Homework 8, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. Let X be the number of distinct birthdays in a group of 110 people (i.e., the number of days in a year such that at least one person in the group has that birthday). Under the usual assumptions (no Feb 29, all the other 365 days of the year are equally likely, and the day when one person is born is independent of the days when the other people are born), find the mean and variance of X.

Let I_j be the indicator r.v. for the event that at least one of the people was born on the jth day of the year, so $X = \sum_{j=1}^{365} I_j$ with $I_j \sim \text{Bern}(p)$, where $p = 1 - (364/365)^{110}$. The I_j's are dependent but by linearity, we still have

$$E(X) = 365p \approx 95.083.$$

By symmetry, the variance is

$$\text{Var}(X) = 365\text{Var}(I_1) + 2\binom{365}{2}\text{Cov}(I_1, I_2).$$

To get the covariance, note that $\text{Cov}(I_1, I_2) = E(I_1 I_2) - E(I_1)E(I_2) = E(I_1 I_2) - p^2$, and $E(I_1 I_2) = P(I_1 I_2 = 1) = P(A_1 \cap A_2)$, where A_j is the event that at least one person was born on the jth day of the year. The probability of the complement is

$$P(A_1^c \cup A_2^c) = P(A_1^c) + P(A_2^c) - P(A_1^c \cap A_2^c) = 2\left(\frac{364}{365}\right)^{110} - \left(\frac{363}{365}\right)^{110},$$

so $\text{Var}(X) = 365p(1-p) + 365 \cdot 364 \cdot (1 - (2\left(\frac{364}{365}\right)^{110} - \left(\frac{363}{365}\right)^{110}) - p^2) \approx 10.019.$

2. A scientist makes two measurements, considered to be independent standard Normal r.v.s. Find the correlation between the larger and smaller of the values.
Hint: note that $\max(x, y) + \min(x, y) = x + y$ and $\max(x, y) - \min(x, y) = |x - y|$.

Let X and Y be i.i.d $\mathcal{N}(0, 1)$ and $M = \max(X, Y), L = \min(X, Y)$. By the hint,

$$E(M) + E(L) = E(M + L) = E(X + Y) = E(X) + E(Y) = 0,$$

$$E(M) - E(L) = E(M - L) = E|X - Y| = \frac{2}{\sqrt{\pi}},$$

where the last equality was shown in class. So $E(M) = 1/\sqrt{\pi}$, and

$$\text{Cov}(M, L) = E(ML) - E(M)E(L) = E(XY) + (EM)^2 = (EM)^2 = \frac{1}{\pi},$$

since $ML = XY$ has mean $E(XY) = E(X)E(Y) = 0$. To obtain the correlation, we also need $\text{Var}(M)$ and $\text{Var}(L)$. By symmetry of the Normal, $(-X, -Y)$ has the same distribution as (X, Y), so $\text{Var}(M) = \text{Var}(L)$; call this v. Then

$$E(X-Y)^2 = \text{Var}(X-Y) = 2, \text{ and also}$$

$$E(X-Y)^2 = E(M-L)^2 = EM^2 + EL^2 - 2E(ML) = 2v + \frac{2}{\pi}.$$

So $v = 1 - \frac{1}{\pi}$ (alternatively, we can get this by taking the variance of both sides of $\max(X, Y) + \min(X, Y) = X + Y$). Thus,

$$\text{Cor}(M, L) = \frac{\text{Cov}(M, L)}{\sqrt{\text{Var}(M)\text{Var}(L)}} = \frac{1/\pi}{1 - 1/\pi} = \frac{1}{\pi - 1}.$$

3. Athletes compete one at a time at the high jump. Let X_j be how high the jth jumper jumped, with $X_1, X_2, \dots$ i.i.d. with a continuous distribution. We say that the jth jumper set a *record* if X_j is greater than all of $X_{j-1}, \dots, X_1$.

(a) Find the variance of the number of records among the first n jumpers (as a sum). What happens to the variance as $n \to \infty$?

Let I_j be the indicator r.v. for the jth jumper setting a record. By symmetry, $E(I_j) = P(I_j = 1) = 1/j$ (as all of the first j jumps are equally likely to be the largest of those jumps). It was shown on HW 4 that I_{110} and I_{111} are independent. Similarly, I_i is independent of I_j for all i, j with $i < j$ (it turns out that they are independent, not just pairwise independent). To see this, note that by symmetry, learning that the jth jumper sets a record gives no information whatsoever about how the first i jumpers rank among themselves, or compute

$$P(I_i = I_j = 1) = \frac{\binom{j-1}{j-i-1}(j-i-1)!(i-1)!}{j!} = \frac{(i-1)!(j-1)!}{i!j!} = \frac{1}{ij} = P(I_1 = 1)P(I_2 = 1),$$

where the numerator corresponds to putting the best of the first j jumps in position j, picking any $j-1+1$ of the remaining jumps to fill positions $i+1$ through $j-1$ and putting them in any order, putting the best of the remaining i jumps in position i, and then putting the remaining $i-1$ jumps in any order.

The variance of I_j is $\text{Var}(I_j) = E(I_j^2) - (EI_j)^2 = \frac{1}{j} - \frac{1}{j^2}$. Since the I_j are pairwise independent (and thus uncorrelated), the variance of $I_1 + \cdots + I_n$ is

$$\sum_{j=1}^{n} \left(\frac{1}{j} - \frac{1}{j^2} \right),$$

which goes to ∞ as $n \to \infty$ since $\sum_{j=1}^{n} \frac{1}{j}$ diverges and $\sum_{j=1}^{n} \frac{1}{j^2}$ converges (to $\pi^2/6$, it turns out).

(b) A *double record* occurs at time j if *both* the jth and $(j-1)$st jumpers set records. Find the mean number of double records among the first n jumpers (simplify fully; it may help to note that $\frac{1}{j(j-1)} = \frac{1}{j-1} - \frac{1}{j}$). What happens to the mean as $n \to \infty$?

Let J_j be the indicator r.v. for a double record occurring at time j, for $2 \leq j \leq n$. Then $P(J_j = 1) = \frac{1}{j(j-1)}$ (again by independence of the record indicators). So the expected number of double records is

$$\sum_{j=2}^{n} \frac{1}{j(j-1)} = \sum_{j=2}^{n} \left(\frac{1}{j-1} - \frac{1}{j} \right) = 1 - \frac{1}{n},$$

since all the other terms cancel out. Thus, the expected number of records goes to ∞, but the expected number of double records goes to 1.

4. Find the PDF of Z^4 for $Z \sim \mathcal{N}(0,1)$.

Let $Y = Z^4$. For $y > 0$, the CDF of Y is

$$P(Y \leq y) = P(Z^4 \leq y) = P(-y^{1/4} \leq Z \leq y^{1/4}) = \Phi(y^{1/4}) - \Phi(-y^{1/4}) = 2\Phi(y^{1/4}) - 1.$$

So the PDF is

$$f_Y(y) = \frac{2}{4} y^{-3/4} \varphi(y^{1/4}) = \frac{1}{2\sqrt{2\pi}} y^{-3/4} e^{-y^{1/2}/2},$$

for $y > 0$, where φ is the $\mathcal{N}(0,1)$ PDF.

5. Let X and Y be independent positive r.v.s, with PDFs f_X and f_Y respectively, and consider the product $T = XY$. When asked to find the PDF of T, Jacobno argues that "it's like a convolution, with a product instead of a sum. To have $T = t$ we need $X = x$ and $Y = t/x$ for some x; that has probability $f_X(x) f_Y(t/x)$, so summing up these possibilities we get that the PDF of T is $\int_0^\infty f_X(x) f_Y(t/x) dx$." Evaluate Jacobno's argument, while getting the PDF of T (as an integral) in 2 ways:

(a) using the continuous law of total probability to get the CDF, and then taking the derivative (you can assume that swapping the derivative and integral is valid);

By the law of total probability (conditioning on X),

$$\begin{aligned}P(T \le t) &= \int_0^\infty P(XY \le t | X = x) f_X(x) dx \\ &= \int_0^\infty P(Y \le t/x | X = x) f_X(x) dx \\ &= \int_0^\infty F_Y(t/x) f_X(x) dx,\end{aligned}$$

which has derivative

$$f_T(t) = \int_0^\infty f_X(x) f_Y(t/x) \frac{dx}{x}.$$

This is *not* the same as Jacobno claimed: there is an extra x in the denominator. This stems from the fact that the transformation (X, Y) to (XY, X) is nonlinear, in contrast to the transformation (X, Y) to $(X + Y, X)$ considered in SP 8 # 2.3. Jacobno is ignoring the distinction between probabilities and probability *densities*, and is implicitly (and incorrectly) assuming that there is no Jacobian term.

(b) by taking the log of both sides of $T = XY$ and doing a convolution (and then converting back to get the PDF of T).

Let $Z = \log(T), W = \log(X), V = \log(Y)$, so $Z = W + V$. The PDF of Z is

$$f_Z(z) = \int_{-\infty}^\infty f_W(w) f_V(z - w) dw,$$

where by change of variables $f_W(w) = f_X(e^w)e^w, f_V(v) = f_Y(e^v)e^v$. So

$$f_Z(z) = \int_{-\infty}^\infty f_X(e^w)e^w f_Y(e^{z-w})e^{z-w} dw = e^z \int_{-\infty}^\infty f_X(e^w) f_Y(e^{z-w}) dw.$$

Transforming back to T, we have

$$f_T(t) = f_Z(\log t)\frac{1}{t} = \int_{-\infty}^\infty f_X(e^w) f_Y(e^{\log(t)-w}) dw = \int_0^\infty f_X(x) f_Y(t/x) \frac{dx}{x},$$

letting $x = e^w$. This concurs with (a): Jacobno is missing the x in the denominator.

6. Let X, Y be continuous r.v.s with a *spherically symmetric* joint distribution, which means that the joint PDF is of the form $f(x, y) = g(x^2 + y^2)$ for some function g. Let (R, θ) be the polar coordinates of (X, Y), so $R^2 = X^2 + Y^2$ is the squared distance from the origin and θ is the angle (in $[0, 2\pi)$), with $X = R\cos\theta, Y = R\sin\theta$.

(a) Explain intuitively why R and θ are independent. Then prove this by finding the joint PDF of (R, θ).

Intuitively, this makes sense since the joint PDF of X, Y at a point (x, y) only depends on the distance from (x, y) to the origin, not on the angle, so knowing R gives no information about θ. The absolute Jacobian is r (as shown on the math review handout), so

$$f_{R,\theta}(r,t) = f_{X,Y}(x,y)r = r \cdot g(r^2)$$

for all $r \geq 0, t \in [0, 2\pi)$. This factors as a function of r times a (constant) function of t, so R and θ are independent with $\theta \sim \text{Unif}(0, 2\pi)$.

(b) What is the joint PDF of (R, θ) when (X, Y) is Uniform in the unit disc $\{(x, y) : x^2 + y^2 \leq 1\}$?

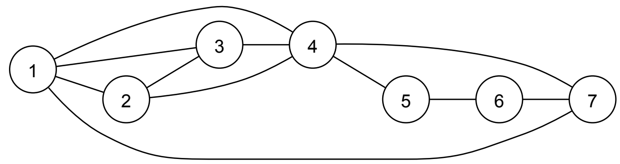
We have $f_{X,Y}(x, y) = \frac{1}{\pi}$ for $x^2 + y^2 \leq 1$, so $f_{R,\theta}(r, t) = \frac{r}{\pi}$ for $0 \leq r \leq 1, t \in [0, 2\pi)$ (and the PDF is 0 otherwise). This says that R and θ are independent with marginal PDFs $f_R(r) = 2r$ for $0 \leq r \leq 1$ and $f_\theta(t) = \frac{1}{2\pi}$ for $0 \leq t < 2\pi$.

(c) What is the joint PDF of (R, θ) when X and Y are i.i.d. $\mathcal{N}(0, 1)$?

The joint PDF of X, Y is $\frac{1}{2\pi}e^{-(x^2+y^2)/2}$, so $g(r^2) = \frac{1}{2\pi}e^{-r^2/2}$ and the joint PDF of (R, θ) is $\frac{1}{2\pi}re^{-r^2/2}$. This says that R and θ are independent with marginal PDFs $f_R(r) = re^{-r^2/2}$ for $r \geq 0$ and $f_\theta(t) = \frac{1}{2\pi}$ for $0 \leq t < 2\pi$. (The distribution of R is called *Weibull*; note that it is the distribution of $W^{1/2}$ for $W \sim \text{Expo}(1/2)$.)

7. A *network* consists of n *nodes*, each pair of which may or may not have an *edge* joining them. For example, a social network can be modeled as a group of n nodes (representing people), where an edge between i and j means they know each other. Assume the network is undirected and does not have edges from a node to itself (for a social network, this says that if i knows j, then j knows i and that, contrary to Socrates' advice, a person does not know himself or herself). A *clique* of size k is a set of k nodes where every node has an edge to every other node (i.e., within the clique, everyone knows everyone). An *anticlique* of size k is a set of k nodes where there are no edges between them (i.e., within the anticlique, no one knows anyone else). For example, the picture below shows a network with nodes labeled $1, 2, \ldots, 7$, where $\{1, 2, 3, 4\}$ is a clique of size 4, and $\{3, 5, 7\}$ is an anticlique of size 3.

(a) Form a random network with n nodes by independently flipping fair coins to decide for each pair $\{x, y\}$ whether there is an edge joining them. Find the expected number of cliques of size k (in terms of n and k).



Order the $\binom{n}{k}$ subsets of people of size k in some way (i.e., give each subset of size k a code number), and let X_i be the indicator. Since $X_1 + X_2 + \cdots + X_{\binom{n}{k}}$ is the number of cliques of size k, the expected number is

$$E(X_1 + X_2 + \cdots + X_{\binom{n}{k}}) = \binom{n}{k} E(X_1) = \binom{n}{k} P(X_1 = 1) = \frac{\binom{n}{k}}{2^{\binom{k}{2}}}.$$

(b) A *triangle* is a clique of size 3. For a random network as in (a), find the variance of the number of triangles (in terms of n).

Hint: find the covariances of the indicator random variables for each possible clique. There are $\binom{n}{3}$ such indicator r.v.s, some pairs of which are independent and other pairs of which are dependent.

Let $k = 3$ and the X_i be as in (a). Then

$$\begin{aligned}\mathrm{Var}(X_1 + \cdots + X_{\binom{n}{3}}) &= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_{\binom{n}{3}}) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j) \\ &= \binom{n}{3} \mathrm{Var}(X_1) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)\end{aligned}$$

We have

$$\mathrm{Var}(X_1) = 2^{-\binom{3}{2}}(1 - 2^{-\binom{3}{2}}) = \frac{7}{64}$$

since a Bern(p) r.v. has variance $p(1-p)$.

To compute $\mathrm{Cov}(X_i, X_j)$ for $i < j$, consider how many people are in common for group i and group j. If the number of people in common is 0 or 1 (as shown in the upper and lower left cases in Figure 1, respectively), then the $\mathrm{Cov}(X_i, X_j) = 0$ since the coin flips used to determine whether Group i is a clique are independent of those used for Group j. If there are 2 people in common (as shown in the lower right case of Figure 1), then

$$\mathrm{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) = \frac{1}{2^5} - (\frac{1}{2^3})^2 = \frac{1}{64},$$

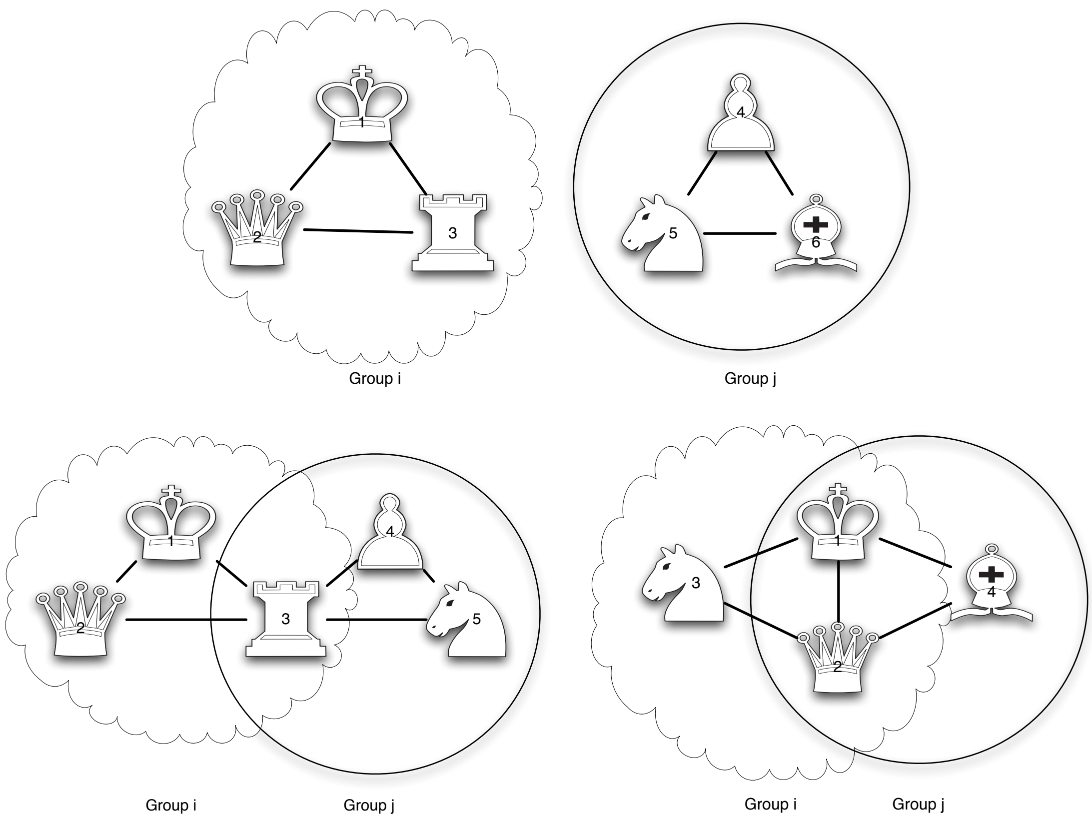


Figure 1: Two groups with 0 (upper), 1 (lower left), 2 (lower right) people in common.

since 5 distinct pairs of people must know each other to make X_iX_j equal to 1.

There are $\binom{n}{4}\binom{4}{2} = 6\binom{n}{4}$ pairs of groups $\{i, j\}$ $(i \neq j)$ with 1 pair of people in common (choose 4 people out of the n, then choose which 2 of the 4 are the overlap of the groups). The remaining pairs of groups have covariance 0. Thus, the variance of the number of cliques is

$$\frac{7}{64}\binom{n}{3} + 2 \cdot 6\binom{n}{4} \cdot \frac{1}{64} = \frac{7}{64}\binom{n}{3} + \frac{3}{16}\binom{n}{4}.$$

(c) Suppose that $\binom{n}{k} < 2^{\binom{k}{2}-1}$. Show that there is a network with n nodes containing no cliques of size k or anticliques of size k.

Hint: explain why it is enough to show that for a random network with n nodes, the probability of the desired property is positive; then consider the complement.

We will prove the existence of a network with the desired property by showing that the probability is positive that a random network has the property is positive. Form a random network as in (a), and let A_i be the event that the ith group of k people (in any fixed ordering) is neither a clique nor an anticlique. We have

$$P\left(\bigcup_{i=1}^{\binom{n}{k}} A_i^c\right) \leq \sum_{i=1}^{\binom{n}{k}} P(A_i^c) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1,$$

which shows that

$$P\left(\bigcap_{i=1}^{\binom{n}{k}} A_i\right) = 1 - P\left(\bigcup_{i=1}^{\binom{n}{k}} A_i^c\right) > 0,$$

as desired. Alternatively, let C be the number of cliques of size k and A be the number of anticliques of size k, and write $C + A = T$. Then

$$E(T) = E(C) + E(A) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1$$

by the method of Part (a). So $P(T = 0) > 0$, since $P(T \geq 1) = 1$ would imply $E(T) \geq 1$. This again shows that there must be a network with the desired property.