1 Joint, marginal, and conditional distributions

1. The arrival times are $X, Y \sim Unif(0,1)$ and we want to compute $P(|X-Y| < .25) = P(-.25 < X - Y < .25)$. Thus we want to integrate the joint distribution of two iid unif(0,1) over a hexagon in the plane (with vertices at $(0,0), (.25,0), (1,.75), (1,1), (.75,1), (0,.25)$). Viewing this hexagon the complement of two isosceles right triangles each of area $(3/4)^2/2 = 9/32$, we get $7/16$ as the probability.

2. (a) There are six equally likely permutations since their arrivals are iid and Carl arrives first in two of then, so $1/3$.
   (b) $A, B, C \sim Unif(0,1)$ where 1 refers to one half hour. Then we want to know $P(A, B > 2/3 | A, B > 1/3) = (1/2)^2 = 1/4$.
   (c) This is $P(AorB > 2/3 | A, B > 1/3) = 1 - P(A, B < 2/3 | A, B > 1/3) = 3/4$.
   (d) This is $P(|A-B| > 1/6 | A, B > 1/3) = P(|X-Y| > 1/6) = (3/6)^2/(4/6)^2 = 9/16$ where $X, Y \sim Unif(1/3, 1)$.

3. (a) $P(X = i, H = 1) = P(X = i | H = 1)P(H = 1) = \binom{n}{i} a^i (1-a)^{n-i} p$ and $P(X = i, H = 0) = P(X = i | H = 0)P(H = 0) = \binom{n}{i} b^i (1-b)^{n-i}(1-p)$.
   (b) $P(X = i) = P(X = i | H = 1)P(H = 1) + P(X = i | H = 0)P(H = 0) = \binom{n}{i} a^i (1-a)^{n-i} p + \binom{n}{i} b^i (1-b)^{n-i}(1-p)$.
   (c) $P(H = 1 | X = k) = P(X = k | H = 1)P(H = 1)/P(X = k) = 1/(1 + (b/a)^i (1-b)^{n-1}/(1-a)^{n-1}(1-p)/p)$.

4. (a)

	$X=0$	$X=1$	$X=2$
$Y=0$	0	1/2	0
$Y=1$	1/4	0	1/4

   (b) $X \sim Bin(2, 1/2)$ and $Y = Dunif(0,1)$.
   (c) They are not independent. Knowing $Y = 1$ means we cannot have $X = 1$.
   (d) $P(Y = 1 | X = x) = 1 - \delta_{x1}$. $P(X = 1 | Y = 0) = 1$ and $P(X = 0 | Y = 1) = P(X = 2 | Y = 1) = 1/2$.

5. (a) Given $X = x$, then $Y \sim Bin(x, p)$. Thus $P(Y = y, X = x) = \binom{x}{y} p^y (1-p)^{x-y}/6$. They are not independent since knowing $X = 1$ means $Y \leq 1$.
   (b) $X \sim Dunif(1,2,3,4,5,6)$ and $P(Y = y) = \sum_{x=1}^{6} \binom{x}{y} p^y (1-p)^{x-y}/6$
   (c) $(Y = y | X = x) \sim Bin(x, p)$. $P(X = x | Y = y) = P(Y = y | X = x)P(X = x)/P(Y = y) = \binom{x}{y} p^y (1-p)^{x-y}/6 / (\sum_{x=1}^{6} \binom{x}{y} p^y (1-p)^{x-y}/6)$.

6. (a) Since $X+Y = r$, then $P(X = x, Y = r-x) = P(X = x)$ and $X \sim HGeom(n, m, r)$.
   (b) Since $X+Y = r$, then $P(X = x, Y = r-x) = P(X = x)$ and $X \sim HGeom(n, m, r)$.
   (c) Given $X = x$, the only possibility is $Y = r - x$, so Y is a constant distribution with $P(Y = r - x) = 1$.

7. (a) $f(x, y) = f(y|x)f(x) = 1/(xL)$ with support $0 \le x \le L$ and $0 \le y \le x$, which defines the triangular region with vertices $(0, 0), (L, 0), (L, L)$.
   (b) $f(x) = \int_0^x f(x, y)dy = \int_0^x 1/xLdx = 1/L$.
   (c) $f(y|x) = f(x, y)/f(x) = 1/(xL)/1/L = 1/x$.
   (d) $f(y) = \int_y^L f(x, y)dx = \log(x)/L|_y^L = \log(L)/L - \log(y)/L$.
   (e) $f(x|y) = f(y|x)f(x)/f(y) = 1/(x\log(L) - x\log(y))$.

8. (a) Let $x + y + z = 5$. Then $P(X = x, Y = y, Z = z) = (\binom{4}{x}\binom{5}{x}\binom{4}{y}\binom{5-x}{y}\binom{44}{z})/(52^5)$.
   (b) Given x, y then z is determined, so $P(X = x, Y = y) = P(X = x, Y = y, Z = 5 - x - y)$.
   (c) $\binom{4}{x}\binom{4}{y}\binom{44}{z}/\binom{52}{5}$.

9. (a) $P(X = x, Y = y, N = x + y) = (1 - p)^{x+y}p^2$.
   (b) $P(X = x, N = n) = P(X = x, Y = n - x, N = n) = (1 - p)^n p^2$.
   (c) N is negative binomial $NBin(2, p)$, so $P(X = x|N = n) = P(X = x, N = n)/P(N = n) = (1 - p)^n p^2/(n + 1)(1 - p)^n p^2 = 1/(n + 1)$. This says that $X \sim Dunif(0, \ldots, n)$ is uniformly distributed given $N = n$.

10. (a) $P(X + Y \le t|X = x) = P(Y \le t - x|X = x) = P(Y \le t - x) = 1 - e^{-\lambda(t-x)}$ for $t \ge x$.
    (b) $f_{T|X}(t|x) = \lambda e^{-\lambda(t-x)}$ for $t \ge x$. This is a valid PDF: $\int_x^\infty \lambda e^{-\lambda(t-x)}dt = \int_0^\infty e^{-\lambda y}dy = 1$ by the exponential PDF.
    (c) $P(T \le t) = P(X + Y \le t) = \int_0^t \lambda e^{-\lambda x}\int_0^{t-x} \lambda e^{-\lambda y}dydx = \int_0^t \lambda e^{-\lambda x} - \lambda e^{-\lambda t}dx = (1 - e^{-\lambda t}) - t\lambda e^{-\lambda t}$. Then the t derivative is $\lambda e^{-\lambda t} - \lambda e^{-\lambda t} + t\lambda^2 e^{-\lambda t} = t\lambda^2 e^{-\lambda t}$. Thus $f_{X|T}(x|t) = f_{T|X}(t|x)f(x)/f(t) = \lambda e^{-\lambda(t-x)}\lambda e^{-\lambda x}/t\lambda^2 e^{-\lambda t} = 1/t$ so $X|T = t \sim Unif(0, t)$.
    (d) Derived this in the previous problem, but $f(t) = f(t|x)f(x)/f(x|t) = t\lambda e^{-\lambda(t-x)}\lambda e^{-\lambda x} = t\lambda^2 e^{-\lambda t}$.

11. (a) $f(x, y, z) = f(y, z|x)f(x) = f(y|x)f(z|x)f(x) = \varphi(y - x)\varphi(z - x)\varphi(x)$.
    (b) They are not unconditionally independent, since knowing Z or Y gives some information about x and hence about the mean of Y or Z.
    (c) $f(y, z) = \int_{-\infty}^\infty f(x, y, z)dx$.

12. (a) $P(X \le x|X > c) = P(c < X < x)/P(c < X) = ((1 - e^{-\lambda x}) - (1 - e^{-\lambda c}))/(1 - (1 - e^{-\lambda c})) = (e^{-\lambda c} - e^{-\lambda x})/e^{-\lambda c} = 1 - e^{-\lambda(x-c)}$. The pdf is $f(x|c) = \lambda e^{-\lambda(x-c)}$.
    (b) $P(X \le x|X < c) = P(X \le \min(c, x))/P(X < c) = (1 - e^{-\lambda\min(x,c)})/(1 - e^{-\lambda c})$. The derivative with respect to x is $(\lambda e^{-\lambda x})/(1 - e^{-\lambda c})$ for $x < c$ and 0 for $x > c$.

13. (a) $P(X \le x|X < Y) = P(X \le x, X < Y)/P(X < Y)$. Since X, Y are iid, $P(X < Y) = 1/2$, so we just need to compute the integral $\int_0^x\int_0^\infty \lambda\lambda e^{-\lambda s}e^{-\lambda t}dsdt = \int_0^x \lambda e^{-\lambda t}e^{-\lambda t}dt = 1/2(1 - e^{-2\lambda x})$ so the CDF is $1 - e^{-2\lambda x}$ from which we derive the PDF $(2\lambda)e^{-(2\lambda)x}$.

(b) Given $X < Y$, we know that $X = \min(X, Y)$ which is distributed as $Expo(2\lambda)$, so the conditional PDF of X given $X < Y$ is $(2\lambda)e^{-2\lambda x}$.

14. (a) The actual length of the stick is irrelevant for this problem, so we let the length be 1. Then the two breakpoints are $X, Y \sim Unif(0, 1)$ and the three lengths are $A = \min(X, Y), B = \max(X, Y) - \min(X, Y) = |X - Y|, C = 1 - \max(X, Y)$. The three lengths can be assembled into a triangle if and only if the sum of the lengths of the two shorter pieces is larger than the length of the longer piece. This is possible if and only if all lengths are less than 1/2. Thus we want $P(\min(X, Y), |X - Y|, 1 - \max(X, Y) < 1/2)$. This corresponds to a subset of the unit square $[0, 1]^2$ (which is the sample space) having area 1/4.

    (b) The actual perimeter of the table is irrelevant for this problem, so we'll assume it's 1. The table can stand as long as all three legs are not placed within a semicircle. The legs are placed in a semicircle if and only if the largest distance between legs is at least 1/2.

    By rotational symmetry, we fix one point (e.g. at $(0, 1/2\pi)$) and then this problem reduces to the same as the previous part where we select two points according to $Unif(0, 1)$ and then need to compute the probability that the largest distance between two points is at most 1/2. Thus the probability is 1/4 that the table will stand.

15. This is by inclusion-exclusion. $P((X, Y) \in [a_1, a_2] \times [b_1, b_2]) = P(X \le a_2, Y \le b_2) - P(X \le a_2, Y \le b_1) - P(X \le a_1, Y \le b_2) + P(X \le a_1, Y \le b_1)$.

16. (a) $\int_0^1 \int_0^1 x + y dx dy = 2 \int_0^1 \int_0^1 x dx dy = 2 \int_0^1 x dx \int_0^1 1 dy = 2(1/2)(1) = 1$.

    (b) The joint PDF does not factor into a function of x and a function of y so X and Y are not independent. (Intuitively, knowing that $X = 1$ for example makes Y twice as likely to be near 1 than to be near 0 while knowing that $X = 0$ for example makes Y infinitely more likely to be near 1 than to be near 0.)

    (c) $f(x) = \int_0^1 x + y dy = (xy + y^2/2)|_0^1 = x + 1/2$ and similarly for $f(y) = y + 1/2$.

    (d) $f(y|x) = f(x, y)/f(x) = (x + y)/(x + 1/2)$.

17. (a) $\int_0^1 \int_0^y cxy dy dx = \int_0^1 cy^3/2 dy = c/8 \implies c = 8$.

    (b) X and Y are not independent because knowing X is large forces Y to be large as well.

    (c) $f(y) = \int_0^y 8xy dx = 4y^3$. $f(x) = \int_0^1 8xy dy = 4x$.

    (d) $f(y|x) = f(x, y)/f(x) = 8xy/4x = 2y$.

18. $f(x, y) = 2$ for $x, y \ge 0, x + 1 \le 1$. $f(x) = \int_0^{1-x} 2dy = 2(1 - x)$ (and similarly for y). $f(x|y) = f(x, y)/f(y) = 2/2(1 - y) = 1/(1 - y)$ (and similarly for y).

19. (a) $f(x, y, z) = 1/(4/3\pi)$ for $x, y, z \in B$ since the volume of the ball is $4/3\pi r^3$ with $r = 1$.

(b) $f(x,y) = \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} 1/(4/3\pi)dz = 2\sqrt{1-x^2-y^2}/(4/3\pi)$

(c) $f(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x,y)dy$.

20. (a) $P(M \le x) = P(Unif(0,1) \le x)^3 = x^3$, so the PDF is $3x^2$. Given $M \le y$, then $U_i \sim Unif(0,y)$ so $P(L \le x|M \le y) = 1 - P(L > x|M \le y) = 1 - (1 - x/y)^3$. Thus $P(L \le x, M \le y) = P(L \le x|M \le y)P(M \le y) = (1-(1-x/y)^3)y^3 = y^3 - (y-x)^3$ and $f(x,y) = d/dxd/dy(y-x)^3 = 6(y-x)$

(b) $f(m|l) = f(l,m)/f(l) = 6(m-l)/3(1-l)^2$.

21. By quadratic formula, we are interested in the probability $P(B^2 > 4A) = \int_0^1 \int_0^{B^2/4} dAdB = 1/12$.

22. (a) The conditional PDF of Y given $X = x$ has support $y > x$.

(b) $f(x,y) = 0$ if $x > y$. However, $f(x), f(y) \neq 0$ for any $0 < x, y$, so $f(x,y) = 0 \neq f(x)f(y)$ so X and Y are not independent.

23. (a) $(U_1, \ldots, U_n)$ is a uniformly random point inside the n-cube $[-1,1]^n$ whose volume is 2^n. The probability that $(U_1, \ldots, U_n)$ is in the unit sphere is the ratio of the volumes of the unit sphere and the unit cube, so $P = v_n/2^n$.

(b) R code: plot$(x, \pi^{x/2}/(\Gamma(x/2+1) * 2^x))$.

(c) $X_n \sim Bin(n, 1-c)$.

(d) If $X_n > 1$, then $P((U_1, \ldots, U_n) \in B_n) = 0$, so $P(X_n > 1) < P((U_1, \ldots, U_n) \notin B_n)$. However, since the probability $p = 1 - c$ is fixed, then $P(X_n > 1) \xrightarrow{n\to\infty} 1 \implies v_n/2^n \xrightarrow{n\to\infty} 0$.

24. (a) $P(Y_1/Y_2 \le y) = P(Y_1 \le yY_2) = \int_0^\infty \int_{y_1/y}^\infty \lambda_1 e^{-\lambda_1 y_1} \lambda_2 e^{-\lambda_2 y_2} dy_2 dy_1 = \frac{\lambda_1}{\lambda_2/y + \lambda_1}$. Intuitively, this makes sense since $yY_2 \sim Expo(\lambda_2/y)$ by the scale transform of the exponential distribution.
$f(y) = \lambda_1\lambda_2/y^2(\lambda_1 + \lambda_2/y)^{-2}$.

(b) A finishes before B if $Y_1 < Y_2 \iff Y_1/Y_2 < 1$. By the previous part, $P(Y_1/Y_2 < 1) = \frac{\lambda_1}{\lambda_2 + \lambda_1}$.

25. (a) Company i crashes when events λ_0 or λ_i happen. Thus $X_i \sim \min(Expo(\lambda_0), Expo(\lambda_i)) \sim Expo(\lambda_0 + \lambda_i)$.

(b) $P(X_1 > x_1, X_2 > x_2) = e^{-\lambda_0 \max(x_1,x_2)} e^{-\lambda_1 x_1} e^{-\lambda_2 x_2}$. Then by inclusion-exclusion, the joint CDF is $P(X_1 < x_1, X_2 < x_2) = 1 - P(X_1 > x_1, X_2 > 0) - P(X_1 > 0, X_2 > x_2) + P(X_1 > x_1, X_2 > x_2)$.

26. (a) $P(Expo(1/10) < Unif(0,10)) = \int_0^{10} \int_0^y 1/10(1/10e^{-1/10x})dxdy = 1/e$.

(b) Fred's waiting time is $L = \min(Unif(0,10), Expo(1/10))$. Then $P(L \le t) = 1 - P(L > t) = 1 - P(Unif(0,10) > t)P(Expo(1/10) > t) = 1 - (10-t)/10 * e^{-t/10}$.

27. (a) $(N_0, N_1, N_2) \sim Mult(3, ((1-p)^2, 2p(1-p), p^2))$.

(b) Since $h = 2N_2 + N_1$, we have $P(N_2 = k|2N_2 + N_1 = h) = P(N_2 = k, N_1 = h - 2k)/P(2N_2 + N_1 = h)$. The denominator can be evaluated as a sum over all possible values of N_2 (which then determines N_1 and N_0).

(c) $P(N_2 = k|2N_2+N_1 = h) = P(N_2 = k, N_1 = h-2k, N_0 = n-h+k)/P(2N_2+N_1 = h) = \frac{\binom{n}{k}\binom{n-k}{h-2k}}{\sum_j \binom{n}{j}\binom{n-k}{h-2j}}$.

(d) p should not appear in the answer to (c) since given h, the probabilities only depend on how we distribute the h hobbits into the n couples, i.e. knowing h eliminates the need to know p.

(e) Let I_j be the indicator variable of whether the j^{th} coupls has both hobbits reaching eleventy-one. Then $P(I_j = 1) = \binom{h}{2}/\binom{2n}{2} = h(h-1)/(2n)(2n-1) \implies E(\sum_j I_j) = h(h-1)/(4n-2)$.

28. (a) $P(X_i = X_j) = \sum_k P(X_i = X_j|X_j = k)P(X_j = k) = \sum_k p(k)^2$.

(b) $P(I = j, X_I = k) = P(X_I = k|I = j)P(I = j) = p(k)/n$. They are independent since the PMF factors and knowing the value of one does not constrain the possible values of the other.

(c) $P(X_I = k) = \sum_j P(I = j, X_I = k) = \sum_j p(k)/n = p(k)$ so X_I has the same marginal distribution as X_1.

(d) $P(X_I = X_J) = P(X_I = X_J|I = J)P(I = J) + P(X_I = X_J|I \neq J)P(I \neq J) = 1/n + (n-1)/n \sum_k p(k)^2$.

29. (a) $P(L = l, M = m) = P(M = m|L = l)P(L = l)$. By the memoryless property of the geometric, $P(M = m|L = l) = P(M = m - l)$ so $P(L = l, M = m) = (1-p)^{m-l}p * (1-p)^{2l}(1-(1-p)^2)$.

(b) The marginal distribution of L is $Geom(1-(1-p)^2)$.

(c) $E(M) = E(M - L) + E(L) = (1-p)/p + (1-p)^2/(1-(1-p)^2)$.

(d) L and $M - L$ are independent by the memoryless property of the geometric, since they are "independent increments." The joint PMF is similar to what was found above with M shifted down to $M - L$.

30. (a) The joint pdf is $f(x, y) = (\theta + (1+\theta x)(1+\theta y))e^{-(x+y+\theta xy)}$. They are independent if $\theta = 0$. Otherwise, the exponential term does not factor as a product of functions of x and y.

(b) If θ is negative, then there are values of x and y where the pdf can be negative.

(c) The marginal PDF is $f(y) = \int_0^\infty f(x, y)dx = 1 - e^{-y}$ and analogously for x, since the antiderivative is $(1+\theta x)e^{-(x+y+\theta xy)}$.

(d) The limit as $y \to \infty$ is $1 - e^{-x}$.

2 2D LOTUS

31. $Var(|X-Y|) = E(|X-Y|^2) - E(|X-Y|)^2 = E((X-Y)^2) - E(\max(X,Y) - \min(X,Y))^2 = Var(X-Y) + E(X-Y)^2 - (2/3-1/3)^2 = Var(X) + Var(Y) + (E(X)-E(Y))^2 - 1/9 = 1/6 + 0 - 1/9 = 1/18 \implies \sigma = 1/3\sqrt{2}$.

32. $E(|X-Y|) = \int_0^\infty \int_0^\infty |x-y|\lambda e^{-\lambda x}\lambda e^{-\lambda y} dydx$. Using the memoryless property, the difference between X and Y can be written as the difference between $\max(X,Y)$ and $\min(X,Y)$ which is distributed as an $Expo(\lambda)$ random variable. Thus $E(\max(X,Y) - \min(X,Y)) = 1/\lambda$.

33. (a) $T = \min(X,Y) \sim Expo(\lambda_1+\lambda_2)$. In terms of double integrals, $E(\min(X,Y)) = \int_0^\infty \int_0^\infty \min(x,y)\lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} dxdy$ and $Var(\min(X,Y)) = \int_0^\infty \int_0^\infty \min(x,y)^2\lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} dxdy$ $E(\min(X,Y))^2$.

    (b) $T \sim Expo(\lambda_1+\lambda_2)$ since the minimum of two exponential random variables is another exponential random variable. Thus $E(T) = \frac{1}{\lambda_2+\lambda_2}$ and $Var(T) = \frac{1}{(\lambda_1+\lambda_2)^2}$.

34. $Cov(X,Y) = E(XY) - E(X)E(Y)$. $E(XY) = \int_0^1 \int_0^{1-x} 2xydydx = \int_0^1 x(1-x)^2 dx = 1/12$. $E(X) = E(Y) = \int_0^1 2x(1-x)dx = 1/3$. Thus $Cov(X,Y) = 1/12 - 1/9 = -1/36$.

35. (a) We can integrate this in polar coordinates: $E(R) = \int_0^{2\pi} \int_0^1 r^2/\pi drd\theta = \int_0^{2\pi} 1/(3\pi)d\theta = 2/3$.

    (b) $P(R \le r) = (\pi r^2)/(\pi) = r^2$. $P(R^2 \le r) = P(R \le \sqrt{r}) = r$ by the previous computation. $f_R(r) = 2r$. $f_{R^2}(r) = 1$. $E(R) = \int_0^1 2r^2 dr 2/3$ and $E(\sqrt{R^2}) = \int_0^1 \sqrt{x}dx = \frac{2}{3}$.

36. (a) $E(X+Y) = \sum_{x,y}(x+y)f(x,y) = \sum_{x,y} xf(x,y) + \sum_{x,y} yf(x,y) = \sum_x x \sum_y f(x,y) + \sum_y y \sum_x f(x,y) = \sum_x xf(x) + \sum_y yg(y) = E(X) + E(Y)$.

    (b) $E(XY) = \sum_{x,y} xyf(x)g(y) = \sum_x xf(x) \sum_y yg(y) = E(X)E(Y)$, where we can only assume that the joint PDF factors since X and Y are independent.

37. (a) $E((X-Y)^2) = \int\int (x-y)^2 f(x,y)dxdy$.

    (b) $\int\int (x^2 - 2xy + y^2)f(x,y)dxdy = \int x^2 \int f(x,y)dxdy + \int y^2 \int f(x,y)dxdy - 2\int xf(x)dx \int yf(y)dy =$ $E(X^2) + E(Y^2) - 2E(X)E(Y) = (\sigma^2+\mu^2) + (\sigma^2+\mu^2) - 2\mu^2 = 2\sigma^2$.

    (c) $E((X-Y)^2) = E(((X-\mu)-(Y-\mu))^2) = E((X-\mu)^2) - 2E(X-\mu)E(Y-\mu) + E((Y-\mu)^2) = \sigma^2 + 0 + \sigma^2 = 2\sigma^2$.

3 Covariance

38. The first statement is correct, but the second is not, since $\max(X,Y)$ and $\min(X,Y)$ do not (typically?) have the same distributions as X and Y.

39. (a) $Cov(X+Y, X-Y) = Cov(X,X) - Cov(X,Y) + Cov(Y,X) - Cov(Y,Y) = Var(X) - Var(Y) = 0$.

(b) No, since if $X+Y=12$, then $X-Y=0$.

40. (a) $Cov(X+Y, X-Y) = Var(X) - Var(Y) = 0$.

    (b) No, since if $X+Y=2$ then $X-Y=0$.

41. For any choice of a, b, c, d, Z And W have mean 0 by linearity of expectation. $Var(Z) = Cov(aX+bY, aX+bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y) = a^2+b^2+2ab\rho$. Similarly, $Var(W) = c^2+d^2+2cd\rho$. Finally, $Cov(Z,W) = ac+bd+(ad+bc)\rho$. There are three equations and four unknowns, so there will generally be infinitely many possibilities, but we can choose e.g. $a=1, b=0, c=-\rho/\sqrt{1-2\rho}, d=1/\sqrt{1-2\rho}$.

42. Let I_j be the indicator of whether day j had a person born on that day. Then $P(I_j=0) = (364/365)^{110}$ so $P(I_j=1) = 1-(364/365)^{110}$. Thus $E(X) = 365(1-(364/365)^{110}) = p$ and $Var(X) = \sum_j Var(I_j) + \sum_{j,k} 2Cov(I_j, I_k) = 365p(1-p) + 365 * 364 * ((1-2(364/365)^{110} - (363/365)^{110}) - p^2)$.

43. (a) $E(XY) = P(XY=1) = P(X=Y=1) = P(X=1)P(Y=1) = E(X)E(Y)$ since X and Y can only take values 0 and 1.

    (b) Let X, Y be arbitrary (independent) and $Z = X \oplus Y$ as studied in a previous exercise.

44. $X = \sum_j I_j$ where the $I_j \sim Geom(p_j = (n-j)/n)$ are independent. Thus $Var(X) = \sum_j Var(I_j) = \sum_j q_j/p_j^2 = \sum_j (j/n)/(n-j)^2/n^2 = \sum_j jn/(n-j)^2$.

45. Since $A+B+C=180$, $Cov(A,B) = Cov(A,(B+C)/2) = Cov(A, 90-A/2) = -1/2Var(A)$ so $\rho = -1/2$.

46. Let I_j be the indicator variable of person j getting their own name. Then $E(I_j) = 1/n$ and $E(I_jI_k) = 1/n(n-1)$ so $Var(\sum_j I_j) = \sum_j Var(I_j) + 2\sum_{j<k} Cov(I_j, I_k) = (n-1)/n + 2/(n^2(n-1))$ and the standard deviation is the square root.

47. (a) Each black ball is equally likely to occur between any two white balls, so there are $w+1$ total positions: before first white ball, between first two white balls, between second two white balls, etc. $I_j = 1$ if the black ball lands in any position before the r^{th} white ball so there are r positions and $w+1$ total so $E(I_j) = r/(w+1)$.

    (b) They are positively correlated. Since all orderings of black/white balls are equally likely, knowing that black ball i occurred before the first r white balls gives more positions for black ball j to land within the first r white balls as well.

    (c) We can imagine placing w white balls and 2 black balls, where all arrangements are equally likely. $I_jI_k = 1$ if black balls j and k occur before the r^{th} white ball, so we can place j and k in any of the first $r+1$ positions. Thus $E(I_jI_k) = \binom{r+1}{2}/\binom{w+2}{2}$.

    (d) $Var(X) = \sum_j Var(I_j) + 2\sum_{j<k} Cov(I_j, I_k) = b * r/(w+1) * (w+1-r)/(w+1) + 2(\binom{r+1}{2}/\binom{w+2}{2} - r^2/(w+1)^2)$.

48. Let I_j be the indicator of athlete j setting a record. Then $E(I_j) = 1/j$. Since I_j, I_k are independent, then $E(I_j I_k) = E(I_j)E(I_k)$ so $Var(\sum_j I_j) = \sum_j Var(I_j) = \sum_j (j-1)/j^2 \xrightarrow{n\to\infty} \infty$.

49. By the chicken-egg story, $X \sim Pois(\lambda p)$ and X and $N - X$ are independent, so $Corr(N, X) = Cov(N, X)/\sqrt{\lambda\lambda p} = \frac{Cov(N-X,X)+Cov(X,X)}{\lambda\sqrt{p}} = \sqrt{p}$.

50. Since $Corr(X_i, X_j)$ does not depend on the scaling of X_i and X_j, then we may assume $Var(X_i) = 1$ for all i by rescaling $X_i = X_i/\sqrt{Var(X_i)}$. Then $Var(X_1 + \cdots + X_n) = \sum_j Var(X_j) + 2\sum_{j<k} Cov(X_j, X_k) = \sum_j Var(X_j) + 2\sum_{j<k} \rho\sqrt{Var(X_j)Var(X_k)}$ is positive, so $n + 2\binom{n}{2}\rho > 0 \implies \rho > -1/(n-1)$.

51. $Var(XY) = E((XY)^2) - E(XY)^2 = E(X^2)E(Y^2) - E(X)^2E(Y)^2 = Var(X)Var(Y) + E(X)^2Var(Y) + E(Y)^2Var(X)$.

52. (a) Let I_j be the indicator of whether student j receives the correct shirt. Since all allocations are equally likely, a given student is equally likely to receive any one of the three shirts, so $E(I_j) = 1/3$ and thus $E(\sum_j I_j) = 3n/3 = n$.

    (b) $P(A_1A_2) = \binom{3n-2}{n-2,n,n}/\binom{3n}{n,n,n} = (n-1)/(9n-3)$ and $P(A_1A_{n+1}) = \binom{3n-2}{n-1,n-1,n}/\binom{3n}{n,n,n} = n/(9n-3)$.

    (c) $Var(X) = \sum_j Var(I_j) + 2\sum_{j<k} Cov(I_j, I_k) = 2n/3 + 2(3\binom{n}{2}Cov(A_1, A_2) + (2n^2 + n^2)Cov(A_1, A_{n+1})) = 2n/3 + 3n(n-1)((n-1)/(9n-3) - 1/9) + 6n^2(n/(9n-3) - 1/9)$.

53. (a) X_n is not independent of Y_n, since if $X_n = n$, then $Y_n = 0$.

    (b) $Cov(X_n, Y_n) = \sum_i Cov(A_i, B_i) + \sum_{i\neq j} Cov(A_i, B_j)$ where $A_i, B_i \in \{-1, 0, +1\}$ represents the step in the x and y directions at step i. Note that A_i and B_i are dependent while A_i and B_j are independent for $i \neq j$, so the sum above reduces to $\sum_i Cov(A_i, B_i) = \sum_i E(A_iB_i) - E(A_i)E(B_i)$. Since the drunken man cannot move in both the x and y direction at the same time, one of A_i and B_i is zero so $A_iB_i = 0$. The possible moves are $(+1, 0), (-1, 0), (0, +1), (0, -1)$, so $P(A_i = 1) = P(A_i = -1) = 1/4$ and $P(A_i = 0) = 1/2$ so $E(A_i) = 1(1/4) + -1(1/4) + 0(1/2) = 0$. Thus $E(A_i) = E(B_i) = 0$ so $\sum_i Cov(A_i, B_i) = 0 \implies Cov(X_n, Y_n) = 0$.

    (c) $E(X) = \sum_i E(A_i) = 0$ and similarly for Y and $R^2 = X^2 + Y^2$ so $E(R^2) = E(X^2) + E(Y^2) = Var(X) + E(X)^2 + Var(Y) + E(Y)^2 = \sum_i Var(A_i) + Var(B_i)$. $Var(A_i) = (-1)^2(1/4) + (1)^2(1/4) + 0^2(1/2) = 1/2$ so $\sum_i Var(A_i) + Var(B_i) = n$.

54. Let $X, Y \sim N(0, 1)$ be the observations and $M = \max(X, Y)$ and $L = \min(X, Y)$. Then $Corr(M, L) = Cov(M, L)/\sqrt{Var(M)Var(L)}$. Since $ML = XY, E(ML) = E(XY) = E(X)E(Y) = 0$ since $Cov(M, L) = -E(M)E(L)$. Thus the correlation is $-E(M)E(L)/\sqrt{Var(M)Var(L)}$ so we need to figure out the distribution of the min and max of two normal random variables.

    Since $\max(X, Y) = -\min(-X, -Y)$ and $-X$ and $-Y$ are iid with X and Y (though not independent), then $E(M) = -E(L)$ and $Var(M) = Var(L)$, so the above reduces to $E(M)^2/Var(M)$.

Using the hint, note that $M = (X+Y+|X-Y|)/2 \implies E(M) = 1/2E(X+Y+|X-Y|) = 1/2E(|X-Y|) = 1/2E(|N(0,2)|) = \sqrt{2}/2E(|N(0,1)|) = 1/\sqrt{\pi}$. Thus $E(M)^2 = 1/\pi$.

Next, using the hint again, $E((M+L)^2) = E((X+Y)^2) = 2$ and on the other hand $E((M+L)^2) = E(M^2+2ML+L^2) = 2E(M^2)+2E(XY) = 2E(M^2) \implies E(M^2) = 1$. Thus $E(M)^2/Var(M) = 1/\pi/(1-1/\pi) = 1/(\pi-1)$

55. (a) V has support $(-1,1)$, so we will compute $P(V \leq t)$ for $t \in (-1,1)$. $P(V \leq t) = P(-(t+1)/2 \leq U \leq (t+1)/2) = 1/2(t+1)/2 * 2 = (t+1)/2$. Thus $f(t) = 1/2$ so V is uniform on $(-1,1)$.

    (b) $Cov(U,V) = E(UV) - E(U)E(V)$. Since $E(U) = 0$, $Cov(U,V) = E(UV) = E(2U|U| - U) = \int_{-1}^{1}(2u|u| - u)/2du$. Since both integrands $2u|u|$ and u are odd on the interval $(-1,1)$ this integral is zero so $Cov(U,V) = 0$. U and V are not independent since knowing U completely determines V.

56. (a) $Cov(X,Y) = Cov(V+W, V+Z) = Var(V) = \lambda$.

    (b) X and Y are not independent since they have positive correlation. Given V, $P(X=x, Y=y|V=v) = P(W=x-v, Z=y-v|V=v) = P(W=x-v|V=v)P(Z=y-v|V=v) = P(X=x|V=v)P(Y=y|V=v)$ so X and Y are conditionally independent.

    (c) $P(X=x, Y=y) = \sum_v P(X=x, Y=y|V=v)P(V=v) = \sum_v P(W=x-v)P(Z=y-v)P(V=v)$.

57. Let I_4, I_3, I_2 be the indicators of whether ships of lengths $4,3,2$ are hit. Then $P(I_j = 1) = p_j = \binom{100-j}{5}/\binom{100}{5}$ Thus $E(\sum_j I_j) = \sum_j p_j$. $E(I_jI_k) - E(I_j)E(I_k) = (1 - p_j - p_k + p_{j+k}) - p_jp_k$.

    $Cov(X,X) = \sum_j p_j(1-p_j) + 2\sum_{j<k} Cov(I_j, I_k) = \sum_i p_i(1-p_i) + 2(1 - p_j - p_k + p_{j+k} - p_jp_k)$.

58. (a) Let I_j be the indicator of whether $(X,Y) = (x_i, y_i)$. Then $X - \overline{x} = \sum_j(x_j - \overline{x})I_j$ and $Y - \overline{y} = \sum_j(y_j - \overline{y})I_j$. Covariance is invariant under translations, so $Cov(X,Y) = Cov(X-\overline{x}, Y-\overline{y}) = E(\sum_{j,k}(x_j-\overline{x})(y_j-\overline{y})I_jI_k) - E(X-\overline{x})E(Y-\overline{y}) = \sum_{j,k}(x_j - \overline{x})(y_k - \overline{y})E(I_jI_k)$. Note that if $j \neq k$, then $I_jI_k = 0$ since exactly I_j is nonzero. Thus the sum is $\sum_j(x_j-\overline{x})(y_j-\overline{y})E(I_j) = 1/n\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$ so $Cov(X,Y) = r$.

    (b) For a pair of points (X,Y) and $(\tilde{X},\tilde{Y})$, the signed area of the rectangle with corners (X,Y) and $(\tilde{X},\tilde{Y})$ is $(X-\tilde{Y})(Y-\tilde{Y})$ regardless of the order between (X,Y) and $(\tilde{X},\tilde{Y})$. Each pair of points is equally likely to be sampled in either order, so by the naive definition $E((X-\tilde{X})(Y-\tilde{Y})) = \frac{1}{n^2}\sum_{j,k}(x_j-x_k)(y_j-y_k) = 2/n^2\sum_{j<k}(x_j-x_k)(y_j-y_k)$, where $\sum_{j<k}(x_j-x_k)(y_j-y_k)$ is the total signed area of the rectangles.

    $E(XY - X\tilde{Y} - \tilde{X}Y + \tilde{X}\tilde{Y}) = E(XY) - E(X\tilde{Y}) + E(\tilde{X}\tilde{Y}) - E(\tilde{X}Y)$. Since X and Y are independent of $\tilde{X}$ and $\tilde{Y}$, then we obtain $= E(XY) - E(X)E(\tilde{Y}) + E(\tilde{X}\tilde{Y}) - E(\tilde{X})E(Y) = E(XY) - E(X)E(Y) + E(\tilde{X}\tilde{Y}) - E(\tilde{X})E(\tilde{Y}) = Cov(X,Y) +$

$Cov(\tilde{X}, \tilde{Y}) = 2Cov(X, Y)$. Thus $Cov(X, Y)$ is $1/n^2$ times the total of the signed areas.

(c) (i) The signed area of a rectangle does not depend on the order of the corners. (ii) Scaling the lengths by a_1 and the heights by a_2 scales all the signed areas by $a_1 a_2$. (iii) The signed areas do not change if the rectangles are translated. (iv) The signed area of the rectangle with side lengths $x, y + z$ is equal to the sum of the signed areas of the rectangles with lengths x, y and x, z.

59. (a) $E(\widehat{\theta}) = w_1 E(\widehat{\theta}_1) + w_2 E(\widehat{\theta}_2) = (w_1 + w_2)\theta = \theta$.

(b) $E((\widehat{\theta} - \theta)^2) = E(\widehat{\theta}^2) - \theta^2 = Var(\theta) = w_1^2 Var(\widehat{\theta}_1) + w_2^2 Var(\widehat{\theta}_2) = w_1^2 Var(\widehat{\theta}_1) + (1 - w_1)^2 Var(\widehat{\theta}_2)$. Taking a derivative with respect to w_1 and setting equal to zero yields $2w_1 Var(\widehat{\theta}_1) = 2(1 - w_1) Var(\widehat{\theta}_2) \implies w_1(2Var(\widehat{\theta}_1) + 2Var(\widehat{\theta}_2)) = 2Var(\widehat{\theta}_2) \implies w_1 = \frac{Var(\widehat{\theta}_2)}{Var(\widehat{\theta}_2) + Var(\widehat{\theta}_1)}$.

(c) In this case, the variance of the sample means is $Var(\widehat{\theta}_1) = \sigma^2/n$ and $Var(\widehat{\theta}_2) = \sigma^2/m$ so the ratio of their variances is $\frac{1/m}{1/m+1/n} = \frac{n}{n+m}$.

4 Chicken-egg

60. (a) Let A be the number of voters for A and B the number of voters for B. Then $V = A - B$ and $E(V) = E(A) - E(B) = p\lambda - (1 - p)\lambda = (2p - 1)\lambda$ since unconditionally $A \sim Pois(\lambda p)$ and $B \sim Pois(\lambda(1 - p))$ by the Chicken-egg story.

(b) $Var(V) = Var(A - B) = Var(A) + Var(B) = \lambda$ since A and B are independent by the Chicken-egg story.

61. (a) $P(X = x, Y = y, N = n)$ is zero unless $x + y = n$, in which case we have $P(X = x, Y = y, N = n) = P(X = x, Y = y|N = n)P(N = n) = P(X = x|N = n)P(N = n) = P(Y = y|N = n)P(N + n)$ since given n, X and Y are completely dependent and distributed as $X \sim Bin(n, p)$ and $Y \sim Bin(n, 1 - p)$ by the Chicken-egg story.

(b) $P(X = x, N = n) = P(X = x|N = n)P(N = n) \neq P(X = x)P(N = n)$ since given $N = n$, $X \sim Bin(n, p)$ but unconditionally we have $X \sim Pois(\lambda p)$ so X and N are not independent.

(c) X and Y are independent Poisson random variables with parameters λp and $\lambda(1 - p)$ respectively, so $P(X = x, Y = y) = e^{-\lambda}(\lambda p)^x/x!(\lambda(1 - p))^y/y!$.

62. Let P and N be the number of people who did and did not make a purchase and $V = P+N$. We are given that the probability someome makes a purchase is $.1/2+.9/3 = .35$. Then by the chicken-egg story, we want to find $P(P = k|N = 42) = P(P = k|N = 42, V = k + 42)P(V = k + 42|N = 42) = P(V = k + 42|N = 42) = P(N = 42|V = k + 42)P(V = k + 42)/P(N = 42) = (\binom{k+42}{42}.65^{42}.35^k e^{-100}(100)^{k+42}/(k + 42)!)/(e^{-65}(65)^{42}/42!) = e^{-35}35^k/k!$ so $P \sim Pois(\lambda * .35)$. (This makes sense, since P

and N are unconditionally independent Poisson random variables, so knowing $N = 42$ does not give any information about P.)

63. $P(X = x) = \sum_m P(X = x|N = m)P(N = m) = \sum_{m=x}^{n} \binom{m}{x} s^x(1-s)^{m-x}\binom{n}{m}p^m(1-p)^{n-m} = s^x/(1-s)^x(1-p)^n \sum_{m=x}^{n}(p(1-s)/(1-p))^m \binom{m}{x}\binom{n}{m}$ (I don't recognize this...)
   $P(X = x, Y = y) = P(X = x, Y = y|N = x+y)P(N = x+y) = P(X = x|N = x+y)P(N = x+y) = \binom{x+y}{x}s^x(1-s)^y\binom{n}{x+y}p^{x+y}(1-p)^{n-x-y}$. This does not factor into a function of x and a function of y so X and Y are not independent.

64. (a) $E(\binom{X}{4}) = E(X(X-1)(X-2)(X-3)/4!) = 1/4!E(X^4 - 6X^3 + 11X^2 - 6X)$. The MGF of X is $E(e^{tX}) = \sum_k e^{-\lambda}(\lambda e^t)^k/k! = e^{-\lambda}e^{e^t\lambda} = e^{\lambda(e^t-1)}$ so the first four moments are $E(X) = \lambda, E(X^2) = \lambda^2 + \lambda, E(X^3) = \lambda^3 + 3\lambda^2 + \lambda, E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$ and thus $E(\binom{X}{4}) = \lambda^4/4!$.
   (b) Given X, X_{am} and X_{pm} are completely dependent binomial random variables. Thus by the chicken-egg story, X_{am} and X_{pm} are i.i.d. $Pois(\lambda/2)$ random variables. (Note they are independent, but not conditionally independent given X). Thus $P(X_{am} = a, X_{pm} = b) = P(X_{am} = a)P(X_{pm} = b) = e^{-\lambda}(\lambda/2)^{a+b}/a!b!$.
   (c) Given X, X_1, X_2, X_3, X_4 are multinonomial $Mult(n, (1/4, 1/4, 1/4, 1/4))$ and unconditionally, each $X_i \sim Pois(\lambda/4)$ is independent. Thus the number of choices of 4 non-conflicting courses is $X_1X_2X_3X_4$ since one course can be chosen from each time slot, so the expected number of choices of 4 non-conflicting courses is $E(X_1)E(X_2)E(X_3)E(X_4) = \lambda^4/4^4$ so the ratio is $4^4/4! = 64/6 = 32/3$. This makes sense since if we do not care about conflicting enrollment, then there are more options.

5 Multinomial

65. Let $I_1, \dots, I_n$ and $J_1, \dots, I_n$ be the indicators of whether balls $1, \dots, n$ are placed in bins i and j respectively. Then $Cov(X_i, X_j) = \sum_{j,k} Cov(I_j, J_k) = \sum_k Cov(I_k, J_k) + \sum_{j \neq k} Cov(I_j, J_k) = \sum_k E(I_kJ_k) - E(I_k)E(J_k) = \sum_k -p_ip_j = -np_ip_j$.

66. Let X_i be the number of people born on the i^{th} day. Then $(X_1, \dots, X_{365}) \sim Mult(100, (1/365, \dots, 1/36$
   so by the previous problem, $Cov(X_1, X_2) = -100/365^2$. All the X_i are iid $Bin(100, 1/365)$, so $Var(X_1) = Var(X_2) = 100 * 364/365^2$. Therefore $Corr(X_1, X_2) = -1/364$.

67. (a) With replacement, $(X, Y, Z) \sim Mult(n, ((a+b)/n, c/n, d/n))$ is distributed as a multionomial vector.
   (b) Without replacement, $P(X = x, Y = y, Z = z) = \binom{a+b}{x}\binom{c}{y}\binom{d}{z}/\binom{a+b+c+d}{n}$ for $x + y + z = n$.

68. (a) $(X, Y, Z) \sim Mult(n, (1/3, 1/3, 1/3))$ is distributed as a multionomial vector.
   (b) The game is decisive if all players collectively choose exactly two of RPS. The probability that everyone chooses at most two of three choices is $2^n/3^n$. The intersection of (everyone choosing at between R,P) and (everyone choosing between

P,S) is (everyone choosing P), so by inclusion exclusion the probability that all players collectively choose at most two is $3 * 2^n/3^n - 3/3^n$. We need to further subtract the cases when all players choose exactly one, which can happen in 3 ways, so th efinal answer is $3 * 2^n/3^n - 3/3^n - 3/3^n = (2^n - 2)/3^{n-1}$.

(c) For $n = 5$, this probability is $30/81$. As $n \to \infty$, this probability goes to zero. This makes sense since the probability of everyone choosing at most two of the choices is approximately $2^n/3^{n-1}$ which converges to zero for large n.

69. (a) The three time intervals are disjoint so X, Y, Z are independent and $X \sim Pois(3\lambda)$ while $Y, Z \sim Pois(6\lambda)$. The joint PMF factors by independence, so $P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z)$.

(b) Given $X + Y + Z = 36$, the three rvs become dependent with joint distribution $(X, Y, Z)|X + Y + Z = 36 \sim Mult(36, (1/5, 2/5, 2/5))$.

(c) Given $X + Y + Z = 36$, the conditional PMF of $X + Y$ is $X + Y \sim Bin(36, 3/5)$.

(d) $E(X + Y|X + Y + Z = 36) = 108/5$ and $Var(X + Y|X + Y + Z = 36) = 216/25$.

70. (a) By the chicken egg story, $Y \sim Pois(\lambda p)$ and $Z \sim Pois(\lambda(1 - p))$ are independent with $Y + Z = X$. Then, $Cov(X, Y) = Cov(Y + Z, Y) = Cov(Y, Y) + Cov(Z, Y) = Var(Y) = \lambda p$, while $Var(X) = \lambda$ and $Var(Y) = \lambda p$ so $Cor(X, Y) = \sqrt{p}$.

(b) Let W be non statistics majors. Then $Y + Z + W = n$ and $(Y, Z, W) \sim Mult(n, (rp, r(1 - p), 1 - r))$. $Y \sim Bin(n, rp), Z \sim Bin(n, r(1 - p)), W \sim Bin(n, 1 - r)$.

(c) Analogously to before, $Cov(X, Y) = Cov(Y + Z, Y) = Var(Y) + Cov(Z, Y)$. Since Z, Y are now part of a multinomial random vector, their covariance is $Cov(Z, Y) = -n(rp)(r(1 - p))$ and $Var(Y) = n(rp)(1 - rp)$ and $Var(X) = nr(1 - r)$. Thus $Cov(X, Y) = n(rp)(1 - rp) - n(rp(r(1 - p))) = nrp(1 - r)$ and $Cor(X, Y) = nrp(1 - r)/\sqrt{n(rp)(1 - rp)nr(1 - r)} = \sqrt{p(1 - r)/(1 - rp)}$.

71. (a) $(X_1, X_2, X_3) \sim Mult(n, (p^2, 2p(1 - p), (1 - p)^2))$.

(b) $X_1 + X_2 \sim Bin(n, p^2 + 2p(1 - p))$.

(c) This number is $A = 2X_1 + X_2$. $P(A = a) = \sum_k P(X_1 = k, X_2 = a - 2k, X_3 = n - a + k)$

(d) $P(X_1 = a, X_2 = b, X_3 = c) = \binom{n}{a,b,c}(p^2)^a(2p(1 - p))^b((1 - p)^2)^c$. We maximize the RHS by taking a logarithmic derivative wrt p: $\frac{d}{dp}\log(-) = \frac{d}{dp}(2a\log(p) + b\log(2p) + b\log(1 - p) + 2c\log(1 - p)) = 2a/p + b/p - b/(1 - p) - 2c/(1 - p) = (2a + b)/p - (b + 2c)/(1 - p)$. Setting equal to zero gives $(2a + b) - p(2a + b) = (b + 2c)p \implies p = \frac{2a+b}{2a+2b+2c}$

(e) $P(X_1 + X_2 = a, X_3 = b) = \binom{n}{a}(p^2 + 2p(1 - p))^a((1 - p)^2)^b$. Using the same process as before: $\frac{d}{dp}(a\log(p^2 + 2p(1 - p)) + 2b\log(1 - p)) = \frac{d}{dp}(a\log(p) + a\log(2 - p) + 2b\log(1 - p)) = a/p - a/(2 - p) - 2b/(1 - p)$. Setting equal to zero, we obtain a quadratic polynomial in p whose roots are $1 \pm \sqrt{b/(a + b)}$ so we should take $p = 1 - \sqrt{b/(a + b)}$.

6 Multivariate Normal

72. (a) This follows immediately from the definition since any linear combination $a(X+Y)+b(X-Y)=(a+b)X+(a-b)Y$ is a linear combination of X and Y which is normal since (X,Y) is BVN.
    (b) $P(X+Y=a,X-Y=b)=P(X=(a+b)/2,Y=(a-b)/2)=1/(2\pi\sqrt{1-\rho^2})\exp(-1/(2(1-\rho^2))(((a+b)/2)^2+((a-b)/2)^2-2\rho(a+b)/2(a-b)/2))$.

73. (a) $c=1/(2\pi)$ by converting the double integral $\int\int f_{X,Y}(x,y)dydx$ into polar coordinates.
    (b) $X,Y\sim N(0,1)$ since $f_{X,Y}(x,y)$ factors as the product of two standard normal PDFs and this also implies that X,Y are independent.
    (c) (X,Y) is BVN since it has the correct format of PDF (i.e. it's the BVN pdf with correlation $\rho=0$). Alternatively, X,Y are independent $N(0,1)$ so (X,Y) is BVN.

74. (a) $f_{X,Y}$ is radially symmetric (i.e. only depends on the distance from the origin) so $\int\int f_{X,Y}(x,y)dxy=\pi$ is exactly half the integral from the previous problem. Thus $c=1/\pi$.
    (b) X and Y are not independent, since $X>0 \iff Y>0$. For $x>0$, $f(x)=1/\pi\int_0^\infty e^{-x^2/2-y^2/2}dy = e^{-x^2/2}/\pi\int_0^\infty e^{-y^2/2}dy = e^{-x^2/2}/\pi\sqrt{\pi/2} = 1/\sqrt{2\pi}e^{-x^2/2}$ and similarly for $x<0$. Thus $X,Y\sim N(0,1)$.
    (c) (X,Y) is not bivariate normal since the PDF is zero when X and Y have opposite signs.

75. $M(t_1,t_2,t_3)=E(e^{t_1(X+2Y)+t_2(3X+4Z)+t_3(5Y+6Z)})=E(e^{X(t_1+3t_2)+Y(2t_1+5t_3)+Z(4t_2+6t_3)})=E(e^{X(t_1+3t_2)})E(e^{Y(2t_1+5t_3)})E(e^{Z(4t_2+6t_3)})=M(t_1+3t_2)M(2t_1+5t_3)M(4t_2+6t_3)$ where $M(t)=e^{t^2/2}$ is the standard normal MGF.

76. (a) This is MVN since any linear combination is a linear combination of X and Y which is a standard normal.
    (b) $X+Y\sim N(0,2)$ and similarly $S(X+Y)\sim N(0,2)$ by symmetry properties of the normal. However, $(X+Y)+S(X+Y)$ is not normal since it is zero with probability 1/2, so this is not MVN.
    (c) $S(aX+bY)$ is normal by symmetry properties of the normal since X,Y have mean 0.

77. Since (X,Y) is BVN, $(Y-cX,X)$ is also BVN, so independence is equivalent to uncorrelatedness for $Y-cX,X$. $Cov(Y-cX,X)=Cov(Y,X)-cVar(X)=\rho\sigma_1\sigma_2-c\sigma_1^2=0 \implies c=\rho\sigma_2/\sigma_1$.

78. (a) Let I_j be the indicator of whether child j is taller than both parents. Then $P(I_j=1)=1/3$ so $E(\sum_j I_j)=2$.
    (b) Let I_j be the indicator of whether child j is taller than their mother. Then $P(I_j=1)=\int_0^\infty\int_0^y f(x,y)dxdy$ where $f(x,y)$ is the BVN pdf for X_1,Y_j with correlation ρ. Then $E(\sum_j I_j)=6P(I_j=1)$.

7 Mixed practice

79. (a) Conditional on the arrival time being less than 3 minutes, the first arrival time is no longer exponentially distributed. Instead, note first that $T > t <=> N_t = 0$ so $P(T \leq t | T \leq 3) = 1 - P(T > t | T \leq 3) = 1 - P(N_t = 0 | N_3 = 1)$ and we compute the latter probability as $P(N_t = 0|N_3 = 1) = P(N_3 = 1|N_t = 0)P(N_t = 0)/P(N_3 = 1) = e^{-\lambda(3-t)}(\lambda(3-t))e^{-\lambda t}/e^{-3\lambda}(3\lambda) = (3-t)/3 \implies P(T \leq t|T \leq 3) = t/3$. This says that the conditional distribution of the first arrival time (given that it is less than 3 minutes) is uniform in those three minutes.

    (b) $N \sim Pois(10\lambda)$ and $B + O = N$ where B is the number of blue cars and O is the number of other cars. By the chicken-egg story, $B \sim Pois(10\lambda b_i)$ and $O \sim Pois(10\lambda(1 - b_i))$ are independent.

80. (a) By the chicken egg story, $X \sim Pois(\lambda ps)$ since the probability that a registered democrat votes is ps.

    (b) Conditionally given $V = v$, then $X \sim Bin(v, ps)$.

    (c) Each registered democrat shows up to vote with probability s, so $X \sim Bin(d, s)$.

    (d) Any n element subset of the $d + r$ people is equally likely to have shown up to vote, so $X \sim HGeom(d, r, n)$.

81. (a) $P(X_1 = a, X_2 = b, X_3 = c, X_4 = d) = \binom{m}{a}\binom{m}{b}\binom{m}{c}\binom{m}{d}/\binom{4m}{n}$ where $a+b+c+d = n$.

    (b) This is not multinomial since the sampling here is without replacement, so the draws are not independent. $P(X_1 = n) = \binom{m}{n}/\binom{4m}{n} \neq 1/4^n$ so X_1 is not marginally distributed as $Bin(n, 1/4)$.

    (c) $Cov(X_1 + X_2 + X_3 + X_4, X_1 + X_2 + X_3 + X_4) = 4Var(X_1) + 6Cov(X_1, X_3) = 0 \implies Cov(X_1, X_3) = -2/3Var(X_1)$ and $X_1 \sim HGeom(m, 3m, n)$ so this is equal to $-1/8(4m - n)/(4m - 1)n$.

82. If Y is independent of X, then $P(Y < X) = \int_0^\infty P(Y < a|X = a)f_X(a)da = \int_0^\infty P(Y < a)\lambda e^{-\lambda a}da$. Integration by parts yields $-P(Y < a)e^{-\lambda a}|_0^\infty + \int_0^\infty e^{-\lambda a}f_Y(a)da = E(e^{(-\lambda)Y})$ if Y is continuous and we obtain the same result from summation by parts if Y is discrete.

83. (a) Let X be the number of detected typos. Then $X = 1/2P + 1/2F$ where $P \sim Bin(n, p)$ and $F \sim Bin(n, f)$.

    (b) By the chicken-egg story, the number of typos that Prue catches is independent of the number that she misses and both have Poisson distributions, so $P \sim Pois(\lambda p)$.

84. (a) $B = A + X$ so $P(B \leq t) = P(A + X \leq t) = \int_0^t \int_0^{t-a} \lambda e^{-\lambda x}\lambda e^{-\lambda a}dxda = 1 - e^{-\lambda t}(1 + t)$.

    (b) $Cov(A, B) = Cov(A, A + X) = Var(A) = 1/\lambda^2$.

    (c) $T = A + \max(X, Y) = A + \min(X, Y) + (\max(X, Y) - \min(X, Y))$. By the exponential story, $\min(X, Y) \sim Expo(2\lambda)$ and $\max(X, Y) - \min(X, Y) \sim Expo(\lambda)$ so $E(T) = 1/\lambda + 1/2\lambda + 1/\lambda = 5/(2\lambda)$.

85. (a) Let A, C, G, T refer to the numbers of A,C,G,T in the sequence. Then $(A, C, G, T) \sim Mult(n, (p_A, p_C, p_G, p_T))$. Thus $Cov(A, C) = -np_Ap_C$.

(b) Given this information, each sequence with these properties is equally likely, so we can view this as there being $\binom{n}{a}$ positions for the A's (all others being distributed arbitrarily) while there are $\binom{n-1}{a-1}$ positions for the A's if the first two must occur together. Thus the probability is $\binom{n-1}{a-1}/\binom{n}{a} = a/n$.

(c) Let I_j be the indicator of whether position j is CAT. Then $P(I_j = 1) = \binom{n-3}{a-1, c-1, g, t-1}/\binom{n}{a, c, g, t} = act/(n(n-1)(n-2))$ so $E(\sum_j I_j) = (n-2)act/(n(n-1)(n-2)) = act/(n(n-1))$.

86. (a) The area under the curve is $\int_1^0 f(1-H)d(1-H)$ (note the integration limits since when $t_0 = a$ is the low endpoint, we have start at $(1, 1)$) where $y = f(x)$ is the implicit function determined by $x = 1-H(t_0)$ and $y = 1-G(t_0)$. This parametric integral is $\int_a^b (1-G(t_0))(-h(t_0))dt_0 = -1+\int_a^b G(t_0)h(t_0)dt_0 = -\int_a^b g(t_0)H(t_0)dt_0$. Thus the area under the curve is $\int_a^b g(t_0)H(t_0)dt_0 = E(H(\text{"}T|D = 1\text{"}))$. This is the probability that a non-diseased person has T value at most "$T|D = 1$" which is the T value of a diseased individual. In other words, this is the probability that a randomly selected non-diseased person has T value lower than a randomly selected diseased person.

(b) If $g = h$, then $G = H$ then there are equal probabilities that the diseased individual has a higher T value than the non-diseased individual. In this case, the ROC curve is the straight line between $(0, 0)$ and $(1, 1)$ so the area is $1/2$.

If there is such a threshold t_0, then $1-G(t_0) \approx 1$ while $1-H(t_0) \approx 0$ so both the sensitivity and specificity are very close to 1, meaning that a diseased individual will almost surely have a higher T value than a non-diseased individual. In this case, so the ROC curve is very close to the rectangle with corners $(0, 0)$ and $(1, 1)$ whose area is very close to 1.

87. (a) $E(J) = \sum_j j/n = (n+1)/2$. $Var(J) = E(J^2)-E(J)^2 = \sum_j j^2/n-(n+1)^2/4 = (n+1)(2n+1)/6-(n+1)^2/4 = (n+1)(n-1)/12$.

(b) We can think of bucketing $Unif(0, n)$ by rounding up, which yields the discrete uniform from $Unif(0, n)$. Thus the mean and variance should be approximately that of the uniform, which are $n/2$ and $n^2/12$ respectively by location scale transform since $nUnif(0, 1) = Unif(0, n)$.

(c) $\sum_j I_j$ is the number of other random variables X_n is larger than, which is exactly one lower than its rank since the rank of the smallest element starts at 1.

$E(R_n) = 1+\sum_j E(I_j) = (n+1)/2$. $Var(R_n) = Var(\sum_j I_j) = Cov(\sum_j I_j, \sum_j I_j) = \sum_j Var(I_j)+2\sum_{j<k} Cov(I_j, I_k) = (n-1)/4+(n-1)(n-2)/12 = (n-1)(n+1)/12$.

(d) Since the X_i are iid continuous random variables, the ranks are discrete uniform.

88. (a) Enumerate all k-subsets of n vertices as $1, \ldots, N = \binom{n}{k}$ and let I_j be the indicator of whether subset j forms a clique. In order to have a clique, every edge must be present, so $P(I_j = 1) = 1/2^{\binom{k}{2}}$. Thus the expected number of cliques of size k is $\binom{n}{k}/2^{\binom{k}{2}}$.

(b) $Var(X) = Cov(\sum_j I_j, \sum_j I_j)$. If two subsets of 3 vertices share no vertices or 1 vertex, then whether or not they form cliques are independent because they deal with distinct edge sets. Otherwise, if two subsets share 2 vertices, then $Cov(I_j, I_k) = E(I_j I_k) - E(I_j)E(I_k) = 1/32 - 1/64 = 1/64$ and if two subsets share 3 vertices, then $Cov(I_j, I_k) = Var(I_j) = 7/64$. The contribution from sharing three vertices is $\sum_j Var(I_j) = \binom{n}{3}7/64$. The contribution from sharing two vertices is $2\binom{n}{2}\binom{n-2}{2}/64$ so the variance is the sum of these two.

(c) Let A_j, B_j be whether subset j forms a clique and anticlique respectively. Then the expected number of anticliques is $E(\sum_j A_j + B_j) = \binom{n}{k}(E(A_j) + E(sB_j)) = \binom{n}{k}(1/2^{\binom{k}{2}} + 1/2^{\binom{k}{2}}) = \binom{n}{k}/2^{\binom{k}{2}-1}$ which is less than 1 by assumption. Thus, there must be some graph with 0 total cliques and anticliques.

89. (a) We may consider each word as having some probability of being word j. Since Shakespeare presumably knows a large number of words, the probability that any particular word is word j is relatively small. Thus a Poisson approximation makes sense here. It furthermore makes sense to assume that in works by the same author, similar linguistic patterns will apply.

(b) $P(B > 0, A = 0) = P(B > 0)P(A = 0) = (1 - e^{-\lambda})e^{-\lambda}$ and expanding the inner $e^{-\lambda}$ as a Taylor series $e^{-\lambda} = \sum_j (-\lambda)^j/j!$ yields the desired result.

(c) $P(B > 0, A = 0) = \int_0^\infty P(B > 0, A = 0|\lambda = x)f_0(x)dx = \int_0^\infty P(B > 0|\lambda = x)P(A = 0|\lambda = x)f_0(x)dx = \int_0^\infty(\sum_{i\geq 1}(-1)^{i+1}x^i/i!)e^{-x}f_0(x)dx$
$= \sum_{i\geq 1}(-1)^{i+1}\int_0^\infty x^i/i!e^{-x}f_0(x)dx = \sum_{i\geq 1}(-1)^{i+1}\int_0^\infty P(A = i|\lambda = x)f_0(x)dx$
$= \sum_{i\geq 1}(-1)^{i+1}P(A = i)$.

(d) Let I_j be the event that word j appears in play B but not in play A. Then $P(I_j = 1) = \sum_{i\geq 1}(-1)^{i+1}P(X_j = i)$. Then the expected number of distinct words appearing in play B but not in play A is $E(\sum_j I_j) = \sum_j E(I_j) = \sum_j \sum_{i\geq 1}(-1)^{i+1}P(X_j = i) = \sum_{i\geq 1}(-1)^{i+1}\sum_j P(X_j = i) = \sum_{i\geq 1}(-1)^{i+1}E(W_i)$ since $W_i = \sum_k J_k$ is the sum of the indicators of whether word k appeared i times in play A and $P(J_k = 1) = P(X_j = i)$.