# 1 Conditional expectation given an event

1. (a) $E(T) = E(T|1)/3 + E(T|2)/3 + E(T|3)/3 = (\mu_1 + \mu_2 + \mu_3)/3$.

   (b) $Var(T) = E(Var(T|J)) + Var(E(T|J)) = (\sigma_1^2 + \sigma_2^2 + \sigma_3^2)/3 + (\mu_1^2 + \mu_2^2 + \mu_3^3)/3 - (\mu_1 + \mu_2 + \mu_3)^2/9$.

2. By direct calculation, we can (re)derive that $X|X + Y = 30 \sim Bin(30, 1/5)$. (More generally, $X|X + Y = n \sim Bin(n, 1/5)$.) Thus $E(X|X + Y = 30) = 6$.

3. Let $WD$ be the number of women with the disease. Given all the information, $WD \sim HGeom(21, 14, 5)$. Note that the value of $p$ is unimportant here, the relevant information is that each person has probability $p$ of having the disease independently, so it is not more likely that men or women get the disease. Thus a priori the 5 diseased people are equally likely to be any subset of the 35 people. Thus $E(WD) = 2$. (This makes sense since women are $14/35 = 2/5$ proportion so we would expect 2 women in any equally likely sample of 5 people.)

4. (a) $E(X|X \geq 1) = \sum_{x\geq 1} xP(X = x|X \geq 1) = \sum_{x\geq 1} xP(X = x)/P(X \geq 1) = 1/(1 - e^{-\lambda}) \sum_{x\geq 1} xP(X = x) = \lambda/(1 - e^{-\lambda})$ by recognizing the sum as $E(X)$.

   (b) $E(X^2|X \geq 1) = 1/(1-e^{-\lambda}) \sum_{x\geq 1} x^2 P(X = x) = (\lambda^2+\lambda)/(1-e^{-\lambda})$ by recognizing the sum as $E(X^2)$.

5. (a) Let $M$ be the final value the gambler obtains. Then $M \sim Dunif(1, \ldots, 20)|M \geq m = Dunif(m, \ldots, 20)$. Let $N$ be the number of rolls the gambler required to roll $m$ or greater. Then $N \sim FS((21 - m)/20)$ and the payoff is $P = M - N$. Thus the expected payoff is $E(M) - E(N) = (10 + m/2) - 20/(21 - m)$.

   (b) The first dervative of this wrt $m$ is $1/2 - 20/(21 - m)^2$ which has an optimum at $m = 21 - \sqrt{40}$.

6. (a) First, $P(X < 1) = \int_0^1 \lambda e^{-\lambda x} dx = 1 - e^{-\lambda}$. Then $E(X|X < 1) = \int_0^1 xf(x)/P(X < 1)dx = 1/(1 - e^{-\lambda}) \int_0^1 x\lambda e^{-\lambda x} dx = 1/(1 - e^{-\lambda})(x(-e^{-\lambda x})|_0^1 + \int_0^1 e^{-\lambda x} dx) = 1/(1 - e^{-\lambda})(-e^{-\lambda} + (-1/\lambda e^{-\lambda x})|_0^1) = 1/(1 - e^{-\lambda})(-\lambda e^{-\lambda}/\lambda - e^{-\lambda}/\lambda + 1/\lambda) = (1 - (\lambda + 1)e^{-\lambda})/\lambda(1 - e^{-\lambda})$

   (b) $1/\lambda = E(X) = E(X|X < 1)P(X < 1) + E(X|X > 1)P(X > 1) = E(X|X < 1)(1 - e^{-\lambda}) + (1 + 1/\lambda)e^{-\lambda}$ so $E(X|X < 1) = (1 - (\lambda + 1)e^{-\lambda})/\lambda(1 - e^{-\lambda})$.

7. We only get a payoff if $V < 4b$. Thus for a fixed price $b$, the expected payoff given $V$ is $V - b$ if $V < 4b$ and 0 otherwise, so the expected payoff is $\int_0^{4b} v - bdv = 8b^2 - 4b^2 = 4b^2$, so the winning strategy is to always bid 1/4 million.

8. Let $M$ be the indicator of whether we select the envelope with more money, and define the three events $A_1, A_2, A_3, A_4$ to be the four events $T < X, Y$, $X < T < Y$, $Y < T < X$, $T < X, Y$. Then $E(M) = \sum_i E(M|A_i)P(A_i)$. Following the hint, it will suffice to show that each $E(M|A_i)$ is at least 1/2 with at least one exceeding 1/2.

   If $T < X, Y$, then we will not switch. In this case, the probability that we chose the envelope with more money to begin with is 1/2 since the information about $T$ gives no information about $X$ and $Y$ and both orders are possible and equally likely since $X$ and $Y$ were generated independently from the same distribution. (The conditional distributions of $X$ and $Y$ given $X, Y > T$ are the same.)

If $X < T < Y$ then we will switch and as a result select the envelope with more money. If $Y < T < X$, then we will not switch and we have initially selected the envelope with more money. If $X, Y < T$, then we will switch. However, again the conditional distributions of $X, Y$ given $X, Y < T$ are the same, so we will select the envelope with more money with probability $1/2$.

9. (a) Given the value $U_1 = u_1$ of the first envelope, we stay with probability $u_1$ and switch with probability $1 - u_1$. In the first case, the probability that we picked the highest value is $u_1$ and in the second case, the probability that we picked the highest value is $(1 - u_1)$. Thus the conditional expectation of picking the highest value envelope is $U_1^2 + (1 - U_1)^2$, so the unconditional expected value (i.e. probability) is $E(U_1^2) + E((1 - U_1)^2) = 2/3$.

   (b) Given the value $U_1 = u_1$ of the first envelope, then the expected payout is $u_1^2 + (1 - u_1)/2$. Thus $E(X|U_1) = U_1^2 + (1 - U_1)/2$. Then the unconditional expected payout is $E(U_1^2) + E(1 - U_1)/2 = (1/12 + 1/4) + 1/4 = 7/12$.

10. (a) $E(V) = \sum_i E(X_i) = n/2$.

    (b) $E(V|X_i) = (n - 1)/2 + X_i$.

    (c) $E(V|X_1, W_1) = X_1 + (n - 1)X_1/2$. Given that bidder 1 wins the auction (and assuming every other bidder submits a big equal to their conditional expectation), then player 1 received the maximum signal, so every other player's signal was smaller than player 1's, i.e. $\max(X_i) = X_1$. Given this information, the other $X_i \sim Unif(0, X_1)$.

11. (a) We can decompose the flips as $FS(p) + FS(1 - p)$ since we first need to wait for heads, then we need to wait for tails. The expected value is $1/p + 1/(1 - p)$.

    (b) Let $X$ be the expected number of flips until HH and $X_H$ the expected number of flips until HH given currently having flipped a head. Then $X = pX_H + (1-p)X + 1$ and $X_H = (1 - p)X + 1$ so solving for $X$ in terms of $p$ yields $X = (1 + p)/(1 - (1 + p)(1 - p)) = 1/p^2 + 1/p$.

    (c) This turns the previously computed expected values into conditional expectations given $p$. Compute: $E(1/p) = \beta(a - 1, b)/\beta(a, b) = (a + b - 1)/(a - 1)$, $E(1/p^2) = \beta(a - 2, b)/\beta(a, b) = (a + b - 1)(a + b - 2)/(a - 1)(a - 2)$, and $E(1/(1 - p)) = \beta(a, b - 1)/\beta(a, b) = (a + b - 1)/(b - 1)$. Thus expected number of flips until $HT$ is $(a + b - 1)/(a - 1) + (a + b - 1)/(b - 1)$ and the expected number of flips until $HH$ is $(a + b - 1)/(a - 1) + (a + b - 1)(a + b - 2)/(a - 1)(a - 2)$.

12. (a) If the first flip is heads, then the length of the first run is $FS(p)$ while if the first flip is tails, then the length of the first run is $FS(1-p)$. Thus the expected length of the first run is $p/p + (1 - p)/(1 - p) = 2$.

(b) If the first flip is heads, then the second run is tails and the length of the second run is $FS(1-p)$ while if the first flip is tails, then the second run is heads and the length of the second run is $FS(p)$. Thus the expected length of the second run is $p/(1-p)+(1-p)/p$.

13. If the first roll is $i$, then we need to find the probability of rolling at least $i$, which has probability $(7-i)/6$. Thus given the first roll being $i$, the number of rolls needed to get at least $i$ again is $FS((7-i)/6)$. Thus the expected number of additional rolls is $1/6(6/6+6/5+6/4+6/3+6/2+6/1) = 1/6+1/5+1/4+1/3+1/2+1 = 2.45$.

14. (a) Let $X$ be the number of rolls needed to get a 1 followed immediately by a 2 and $X_1$ be the number of rolls needed to get a 2 assuming the current roll is a 1. Then $X = X_1/6+5X/6+1 \implies X = X_1+6$ and $X_1 = X_1/6+4X/6+1 \implies X_1 = 4X/5+6/5$. Then $X = (4X/5+6/5)+6 \implies X/5 = 36/5 \implies X = 36$.

    (b) Let $X$ be the number of rolls to get two consecutive ones and let $X_1$ be the number of rolls to get two consecutive ones if a one was just rolled. Then $X = X_1/6 + 5X/6+1 \implies X = X_1+6$ and $X_1 = 5X/6+1$. Then $X = (5X/6+1)+6 \implies X = 42$.

    (c) $a_1 = 1$ since the first roll always yields "the same value 1 time in a row." Otherwise, in order to roll same $n+1$ times in a row, first roll same $n$ times in a row. Then either we match the current streak with probability 1/6 or start over (another $a_{n+1}$ rolls) with probability 5/6. Thus $a_{n+1} = a_n + 1/6 + 5/6(a_{n+1}) \implies a_{n+1} = 6a_n + 1$.

    (d) By induction (repeatedly applying the previous formula), $a_{n+1} = 6^n + 6^{n-1} + \cdots + 6 + 1 = (6^{n+1}-1)/5$.

# 2 Conditional expectation given a random variable

15. $2\overline{X} = E(X_1+X_2|\overline{X}) = E((w_1+w_2)X_1+(w_1+w_2)X_2|\overline{X}) = E(w_1X_1+w_2X_2|\overline{X}) + E(w_2X_1+w_1X_2|\overline{X}) = 2E(w_1X_1+w_2X_2|\overline{X}) \implies E(w_1X_1+w_2X_2|\overline{X}) = \overline{X}$.

16. $E(S_k|S_n) = \sum_{i=1}^k E(X_i|S_n) = kS_n/n$.

17. Let $I_j$ be the indicator of the $j^{th}$ roommate pair both taking the course. Then $E(I_j) = p^2$ so $E(X) = np^2$. Let $Y$ be the number of roommates pairs where at least one roommate takes the course.

    Given $N$, $E(X|N) = \sum_j E(I_j|N) = \sum_j \binom{2n-2}{N-2}/\binom{2n}{N} = nN(N-1)/(2n(2n-1))$. As a sanity check, note that as guaranteed by Adam's law: $E(E(X|N)) = p^2n = E(X)$.

18. $E((Y-E(Y|X))^2|X) = E(Y^2-2YE(Y|X)+E(Y|X)^2|X) = E(Y^2|X)-2E(Y|X)E(Y|X)+E(Y|X)^2 = E(Y^2|X)-E(Y|X)^2$, where we used the fact that $E(Y|X)$ is known given $X$ so we can take it out.

19. (a) Intuitively, it seems like we should solve $x = (y-b)/a$.

(b) Since $X, Y$ are bivariate normal, any linear combination is bivariate normal so independence of any linear combinations is equivalent to uncorrelatedness. $Cov(Y - cX, X) = Cov(Y, X) - cVar(X) = \rho - c$, so set $c = \rho$ in order to make $V$ and $X$ uncorrelated. Then choose $V = Y - \rho X$.

(c) Setting $W = X - dY$, compute as before $Cov(Y, X - dY) = Cov(Y, X) - dVar(Y) = \rho - d$ so set $d = \rho$ to make $W$ and $Y$ uncorrelated. Then choose $W = X - dY$.

(d) $E(Y|X) = E(\rho X + V|X) = \rho X + E(V) = \rho X$. Similarly, $E(X|Y) = E(\rho Y + W|Y) = \rho Y + E(W) = \rho Y$.

(e) The previous result implies that we should use the same slope for predicting $X$ from $Y$ as we do for predicting $Y$ from $X$. The point is that $X$ and $Y$ are random, so the best measure of predicting how $Y$ changes in response to a change in $X$ is the covariance between $X$ and $Y$ (divided by the variance of $X$). Similarly, the best measure of predicting how $X$ changes in response to a change in $Y$ is the covariance between $Y$ and $X$ (divided by the variance of $Y$). But covariance is symmetric, so in this case the correlation is the best predictor.

In particular, by previous computations, the product of the slopes for the best line for predicting $Y$ from $X$ and for predicting $X$ from $Y$ is $Cov(X, Y)^2/Var(X)Var(Y) = \rho^2$, so the slopes satisfy an inverse relationship but their product is equal to $\rho^2$ rather than 1. Furthermore, when $X$ and $Y$ are standardized, then actually $b_1 = b_2 = \rho$.

20. (a) $X_1|X_2 = x \sim Bin(n - x, p_1/(p_1 + p_3 + p_4 + p_5))$ so $E(X_1|X_2) = (n - X_2)p_1/(p_1 + p_3 + p_4 + p_5)$ and $Var(X_1|X_2) = (n - X_2)p_1/(p_1 + p_3 + p_4 + p_5)(1 - p_1)/(p_1 + p_3 + p_4 + p_5)$. Note that we don't need to use eve's law since we know the conditional distribution.

(b) $X_1|X_2 + X_3 = x \sim Bin(n - x, p_1/(p_1 + p_4 + p_5))$ so $E(X_1|X_2 + X_3) = (n - X_2 - X_3)p_1/(p_1 + p_4 + p_5)$.

21. (a) $E(Y) = E(Y|A)P(A) + E(Y|A^c)P(A^c)$. $E(Y|I_A = 1) = E(Y|A)$ and $E(Y|I_A = 0) = E(Y|A^c)$, so $E(Y|I_A)$ is a random variable which takes on the value $E(Y|A)$ with probability $P(A)$ and $E(Y|A^c)$ with probability $P(A^c)$.

(b) $E(Y|A) = \sum_y P(Y = y|A) = \sum_y P(Y = y, A)/P(A) = 1/P(A) \sum_y yP(Y = y, A)$. If $Y = y$ and $A$ cannot occur simultaneously, then the probability is zero, so the sum only contains those terms for which $Y = y$ and $A$ can occur simultaneously. In particular, the sum is equal to $1/P(A)E(YI_A)$.

Alternatively, $P(YI_A = y) = P(Y = y, I_A = 1) = P(Y = y|A)P(A)$ so $E(YI_A) = \sum_y P(Y = y|A)P(A) = P(A)E(Y|A)$.

(c) $Y = YI_A + YI_{A^c}$ so $E(Y) = E(YI_A) + E(YI_{A^c}) = E(Y|A)P(A) + E(Y|A^c)P(A^c)$.

22. $E(I_A) = E(E(I_A|X))$. $E(I_A|X = x) = P(A|X = x)$, so $E(I_A|X)$ is the random variable with value $P(A|X = x)$. Thus $E(E(I_A|X)) = \int_{-\infty}^{\infty} P(A|X = x)f_X(x)dx$.

23. (a) $E(W) = E(Y) - E(E(Y|X)) = 0$ by Adam's law. $E(W|X) = E(Y|X) - E(E(Y|X)|X) = 0$ by taking out what's known since $E(Y|X)$ is a function of $X$.

    (b) $Var(W) = E(Var(W|X)) + Var(E(W|X)) = E(Var(W|X)) = E(X^2) = 1$.

24. $E(X) = E(X|p_1)P(p_1) + E(X|p_2)P(p_2) = 1/2np_1 + 1/2np_2 = n(p_1+p_2)/2$. $Var(X) = E(Var(X|J)) + Var(E(X|J)) = E(np_J(1-p_J)) + Var(np_J) = n/2(p_1+p_2-p_1^2-p_2^2) + n^2((p_1^2+p_2^2)/2 - (p_1+p_2)^2/4)$.

25. $E(X_{j+1}|X_j) = p(1+f)X_j + (1-p)(1-f)X_j = X_j(1+f(2p-1))$. $E(X_n) = E(E(X_n|X_{n-1})) = (1+f(2p-1))E(X_{n-1}) = \cdots = (1+f(2p-1))^n E(X_0) = (1+f(2p-1))^n x_0$.

    To see where the Kelly criterion comes from, consider maximizing the log instead. $E(\log(X_{j+1})|\log(X_j)) = p\log((1+f)X_j) + (1-p)\log((1-f)X_j) = (p\log(1+f) + (1-p)\log(1-f)) + \log(X_j)$. Thus $E(\log(X_n)) = n(p\log(1+f) + (1-p)\log(1-f)) + \log(x_0)$ which is maximized at $p/(1+f) - (1-p)/(1-f) = 0 \iff p(1-f) = (1-p)(1+f) \iff p - pf = 1 - p + f - fp \iff f = 2p-1$.

26. $X = X_1 + \cdots + X_N$ where $X_i \sim Pois(\lambda_2)$. Then $E(X) = E(E(X|N)) = E(N\lambda_2) = \lambda_1\lambda_2$ and $Var(X) = E(Var(X|N)) + Var(E(X|N)) = E(N\lambda_2) + Var(N\lambda_2) = 2\lambda_1\lambda_2^2$.

27. (a) Given $N \geq 1$, the $N$ arrival times are $Unif(0,240)$, so the first arrival time is the min of $N$ iid $Unif(0,240)$. By scale transform, since the min of $N$ iid $Unif(0,1)$ is $1/(N+1)$, the expected first arrival time is $240/(N+1)$.

    Now we need to compute $E(240/(N+1)|N \geq 1) = 240/(1-e^{-\lambda})\sum_{n\geq 1} 1/(n+1)e^{-\lambda}\lambda^n/n! = 240/(\lambda(e^\lambda - 1))\sum_{n\geq 1}\lambda^{n+1}/(n+1)! = 240/\lambda(e^\lambda - 1 - \lambda)/(e^\lambda - 1)$.

    If $\lambda = 20$, then the above is approximately 12 so the expected time at which the first person arrives given that at least one person shows up is 12 minutes, so 8:12.

    (b) By symmetry, the arrival time of the last person is $240(1 - 1/(N+1))$ which has expected value $240 - 240/\lambda(e^\lambda - 1 - \lambda)/(e^\lambda - 1)$ (which is also symmetric with before!). Thus, the expected last arrival time is 11:48.

28. (a) $E((\widehat{\theta} - \theta)^2) = E(\widehat{\theta}^2) - 2E(\widehat{\theta}\theta) + E(\theta^2)$. Using Adam's law, the middle term is $-2E(E(\widehat{\theta}\theta|\theta)) = -2E(\theta E(\widehat{\theta}|\theta)) = -2E(\theta^2)$ so the overall expected squared difference is $E(\widehat{\theta}^2) - E(\theta^2)$.

    (b) $E((\widehat{\theta} - \theta)^2) = E(\widehat{\theta}^2) - 2E(\widehat{\theta}\theta) + E(\theta^2)$. Using Adam's law, the middle term is $-2E(E(\widehat{\theta}\theta|X)) = -2E(\widehat{\theta}E(\theta|X)) = -2E(\widehat{\theta}^2)$ so the overall expected squared difference is $E(\theta^2) - E(\widehat{\theta}^2)$.

    (c) Suppose $\theta$ is unbiased and the Bayes procedure. Then from above we have that $2E((\widehat{\theta} - \theta)^2) = E(\widehat{\theta}^2) - E(\theta^2) + E(\theta^2) - E(\widehat{\theta}^2) = 0$ so $E((\widehat{\theta} - \theta)^2) = 0$. Since $(\widehat{\theta} - \theta)^2 \geq 0$, the only way for the expected value to be zero is if $\widehat{\theta} - \theta = 0 \iff \widehat{\theta} = \theta$.

29. $Cov(X,Y) = E(XY) - E(X)E(Y) = E(E(XY|X)) - E(X)E(E(Y|X)) = E(Xc) - E(X)E(c) = 0$.

30. Let $X = Z^2$ and $Y = Z$ for $Z$ standard normal. Then $E(XY) - E(X)E(Y) = E(Z^3) - E(Z)E(Z^2) = 0 - 0$ since the odd moments of the standard normal are zero so $X$ and $Y$ are uncorrelated. However, $E(Z|Z^2) = \sqrt{Z^2}$ with probability 1/2 and $-\sqrt{Z^2}$ with probability 1/2.

31. (a) $X = T_Y = \sum_i T_{Y-i} - T_{Y-i-1}$ where $T_0 = 0$ and $Y \sim FS(p)$. Then $E(X) = E(E(X|Y)) = E(E(T_Y|Y)) = E(E(\sum_i T_{Y-i} - T_{Y-i-1}|Y)) = E(Y/\lambda) = 1/p1/\lambda$. $Var(X) = E(Var(X|Y)) + Var(E(X|Y)) = E(Y/\lambda^2) + Var(Y/\lambda) = 1/p1/\lambda^2 + 1/\lambda^2(1-p)/p^2 = 1/p^2 1/\lambda^2$.

    (b) $E(e^{tX}) = E(E(e^{tX}|Y)) = E(E(e^{tT_1})^Y) = E((\lambda/(\lambda-t))^Y) = \sum_y (\lambda/(\lambda-t))^y (1-p)^{y-1} p = p\lambda/(\lambda-t)\sum_y ((1-p)\lambda/(\lambda-t))^y = p\lambda/(\lambda-t)/1-(1-p)\lambda/(\lambda-t) = \lambda p/(\lambda p - t)$. Thus $X \sim Expo(\lambda p)$ since they have the same MGF.

32. (a) Let $B$ be the event of making a purchase and $X$ the amount that a random customer spends. Then $E(X) = E(X|B)P(B) + E(X|B^c)P(B^c) = \mu p$ and $Var(X) = E(Var(X|I)) + Var(E(X|I)) = p\sigma^2 + \mu^2 p(1-p)$.

    (b) The revenue the store makes is $X_1 + \cdots + X_N$ where $N \sim Pois(8\lambda)$ is the number of customers over the 8 hour period and $X_i$ is the amount of money that the $i^{th}$ customer spends. Then $E(X_1 + \cdots + X_N) = E(E(\sum_i X_i|N)) = E(N\mu p) = 8\lambda\mu p$ and $Var(X) = E(Var(X|N)) + Var(E(X|N)) = E(N(p\sigma^2 + \mu p(1-p))) + Var(N\mu p) = 8\lambda(p\sigma^2 + \mu^2 p(1-p)) + 8\lambda\mu^2 p^2$.

    (c) If $X$ is the number of customers who make a purchase and $Y$ is the number of customers who do not make a purchase, then $X + Y = N$ and given $N$, $X \sim Bin(n,p)$ and $Y \sim Bin(n, 1-p)$. Thus by the chicken-egg story, unconditionally $X \sim Pois(8\lambda p)$ and $Y \sim Pois(8\lambda(1-p))$ are independent. Then the revenue is $B_1 + \cdots + B_X$ where $B_i$ is how much the $i^{th}$ person spends (since we know that they made a purchase), so $E(B) = E(E(\sum_i B_i|X)) = E(X\mu) = 8\lambda p\mu$ and the variance is $Var(B) = E(Var(B|X)) + Var(E(B|X)) = E(X\sigma^2) + Var(X\mu) = 8\lambda p\sigma^2 + \mu^2 8\lambda p$.

33. Let $X$ be the expected amount of time until Fred buys a new computer. Then $X = 1/\lambda + pX \implies X = 1/(1-p)\lambda$.

34. (a) Student A's argument is convincing in the sense that they are correctly applying Adam's law, btu they forgot to condition on all the information.

    (b) Student B's argument seems to be fully correct.

    (c) The number of "failed" rolls until a 1 has the same distribution as the number of "failed" rolls until a 6, both being $T \sim Geom(1/6)$. Let $I_1, \ldots, I_T \sim Dunif(1, \ldots, 6)$. Then $T_1 = 1 + I_1 + \cdots + I_T$ and $T_6 = 6 + I_1 + \cdots + I_T$. Thus $E(T_1) = E(E(T_1|T)) = E(1 + 3.5T) = 1 + 3.5 * 5$ while $E(T_6) = E(E(T_6|T)) = E(6 + 3.5T) = 6 + 3.5 * 5$.

35. (a) Given $N$, then $T \sim Bin(N, p)$. Then $E(T) = E(E(T|N)) = E(Np) = p(1-s)/s$ and $Var(T) = E(Var(T|N)) + Var(E(T|N)) = E(Np(1-p)) + Var(Np) = p(1-p)(1-s)/s + p^2(1-s)/s^2$.

    (b) $E(e^{tT}) = E(E(e^{tT}|N)) = E((pe^t + 1 - p)^N) = s/(1 - (pe^t + 1 - p)(1-s))$. This is geometric...

36. $E(e^{tY}) = E(E(e^{tY}|\lambda)) = E(e^{\lambda(e^t-1)}) = (b_0)^{r_0}/(b_0 - e^t + 1)^{r_0} = (b_0/(b_0+1))/(1 - e^t/(b_0 + 1))$ which is exactly the desired MGF.

37. (a) Let $J$ be the index of which $X_J$ $X_j^*$ is equal to. Then $E(X_j^*) = E(E(X_j^*|J))$. $E(X_j^*|J)$ is the random variable which has value $\mu$ with probability 1 so $E(X_j^*) = \mu$.
    $Var(X_j^*) = E(Var(X_j^*|J)) + Var(E(X_j^*|J))$. $Var(X_j^*|J)$ is the random variable which has value $\sigma^2$ with probability 1 so $E(Var(X_j^*|J)) = \sigma^2$. $E(X_j^!J)$ is constant so its variance is zero, so $E(X_j^*) = \mu$ and $Var(X_j^*) = \sigma^2$.

    (b) $E(\overline{X}^*|X_1, \ldots, X_n) = E(X_1^*|X_1, \ldots, X_n) = \overline{X}$. $Var(\overline{X}^*|X_1, \ldots, X_n) = 1/n((X_1^2 + \cdots + X_n^2)/n - n\overline{X}^2/n) = 1/n(1/n \sum_i X_i^2 - \overline{X}^2)$.

    (c) $E(\overline{X}^*) = E(\overline{X}) = \mu$. $Var(\overline{X}^*) = 1/n^2 \sum_i E(X_i^2) - E(\overline{X}^2) + Var(\overline{X}) = 1/n(Var(X_1) + E(X_1)^2 - Var(\overline{X}) - E(\overline{X})^2) + Var(\overline{X}) = (n-1)/n^2\sigma^2 + n\sigma^2/n^2 = (2n-1)\sigma^2/n^2$.

    (d) In Eve's law, the term $Var(E(\overline{X}^*|X_1, \ldots, X_n))$ is equal to the variance of the sample mean of $X_1, \ldots, X_n$ so the variance of the bootstrap sample mean must be larger. Intuitively, this term represents the between-group variance, which is exactly the variance of the sample mean. On the other hand, we have additional variance coming from the randomness used to create the bootstrap sample, which contributes within-group variance as well.

38. (a) $E(C) = E(I_1T_1) + E(I_2T_2) = \mu_1p_1 + \mu_2p_2$.

    (b) $Var(C) = Var(I_1T_1) + Var(I_2T_2) + 2Cov(I_1T_1, I_2T_2) = E(I_1T_1^2) - E(I_1T_1)^2 + E(I_2T_2^2) - E(I_2T_2)^2 + 2E(I_1I_2)E(T_1)E(T_2) - E(I_1T_1)E(I_2T_2) = p_1(\sigma_1^2 + \mu_1^2) - \mu_1^2p_1^2 + p_2(\sigma_2^2 + \mu_2^2) - \mu_2^2p_2^2 + 2p_{12}\mu_1\mu_2 - \mu_1\mu_2p_1p_2$.

39. (a) $Var(X) = E(Var(X|I)) + Var(E(X|I)) = (p\sigma_1^2 + (1-p)\sigma_2^2) + 0$.
    $Var(X) = Var(I_1X_1) + Var(I_2X_2) + 2Cov(I_1X_1, I_2X_2) = E(I_1X_1^2) - E(I_1X_1)^2 + E(I_2X_2^2) - E(I_2X_2)^2 + 2(E(I_1I_2X_1X_2) - E(I_1X_1)E(I_2X_2)) = p\sigma_1^2 + (1-p)\sigma_2^2$.

    (b) $E(X) = 0$. $E(X^4) = E(X^4|I_1 = 1)p + E(X^4|I_2 = 1)q = E(X_1^4)p + E(X_2^4)q = 3p\sigma_1^4 + 3q\sigma_2^4$. Thus the kurtosis is $(3p\sigma_1^4 + 3q\sigma_2^4)/(p\sigma_1^2 + q\sigma_2^2)^2 - 3$.

40. $Var(B) = E(Var(B|X_1)) + Var(E(B|X_1))$. $E(B|X_1) = E(E(Y|X_1, X_2)|X_1) = E(Y|X_1)$ so the previous simplifies to $E(Var(B|X_1)) + Var(A)$. Thus $Var(B) \geq Var(A)$ since $Var(B|X_1) \geq 0$ so the same is true for its expected value.

41. $E(Y|E(Y|X)) = E(E(Y|X, E(Y|X))|E(Y|X)) = E(E(Y|X)|E(Y|X)) = E(Y|X)$.

42. Want to show that $P(Z=1|X,S)=P(Z_1|S)$. $P(Z_1|X,S)=E(Z|X,S)$. By the previous problem, $E(Z|S)=E(Z|E(Z|X))=E(Z|X)$ and since knowing $X$ determines $S=E(Z|X)$ (which is a function of $X$), then $E(Z|X)=E(Z|X,S)$.

43. (a) $Cov(X,Y|Z)=E(XY-XE(Y|Z)-YE(X|Z)+E(X|Z)E(Y|Z)|Z)=E(XY|Z)-E(Y|Z)E(X|Z)$ since $E(X|Z),E(Y|Z)$ are both functions of $Z$.

    (b) $E(E(XY|Z))-E(E(X|Z)E(Y|Z))+E(E(X|Z)E(Y|Z))-E(E(X|Z))E(E(Y|Z))=E(XY)-E(X)E(Y)$.

# 3 Mixed practice

44. (a) Let $N_i$ be the number of times friend $i$ pays for dinner in $k$ dinners. Then $(N_1,\dots,N_n)\sim Mult(k,(1/n,\dots,1/n))$ and the requested probability is $P(\max(N_i)\leq 1)=\binom{n}{k}P(N_1=\dots=N_k=1,N_{k+1}=\dots=N_n=0)$. If $k>n$, then by the pigeonhole principle someone must pay at least twice, so the probability is zero.

    (b) This is the coupon collector problem. The first dinner, one new person pays, so $FS(1)$. Afterwards, a new person pays with probability $n-1/n$ so $FS(n-1/n)$. After two people have paid, a new person pays with probability $n-2/n$ so $FS(n-2/n)$, etc. Thus The number of dinners until everyone has paid at least once is $FS(n/n)+FS(n-1/n)+\dots+FS(1/n)$ with expected values $n+n/2+n/3+\dots+n/n=n(1+1/2+1/3+\dots+1/n)$.

    (c) $A$ and $B$ are entries in a multinomial $Mult(k,(1/n,\dots,1/n))$ vector, so their covariance is $-k/n^2$, which we could also derive from first principles using indicator variables.

45. (a) By properties of the multinomial distribution, given the number $N$ of dinners, the number of dinners Alice pays for is $Bin(N,1/n)$ and the number of free dinners for Alice is $Bin(N,(n-1)/n)$. Thus by the chicken-egg story, the number of dinners Alice pays for is $Pois(\lambda/n)$ and the number of free dinners for Alice is $Pois(\lambda(n-1)/n)$ and these are independent, hence uncorrelated.

    (b) If $C_i$ is the cost of the $i^{th}$ dinner, then the total cost $C$ is $C=C_1+\dots+C_N$. Then $E(C)=E(E(C|N))=E(Na/b)=\lambda a/b$ since the mean of $Gamma(a,b)$ is $a/b$.

        $Var(C)=E(Var(C|N))+Var(E(C|N))=E(Na/b^2)+Var(Na/b)=\lambda a/b^2+a^2/b^2\lambda=\lambda a/b^2(a+1)$ since the variance of $Gamma(a,b)$ is $a/b^2$.

46. (a) If $C_i$ is the number of pages of book $i$ and $C=C_1+\dots+C_N$ is the total number of pages, then $E(C)=E(E(C|N))=E(N\mu)=\lambda\mu$.

    (b) $Var(C)=E(Var(C|N))+Var(E(C|N))=E(N\lambda)+Var(N\mu)=\lambda\mu+\lambda\mu^2$.

    (c) By the chicken-egg story, the number of books Joe likes is independent of the number of books he dislikes if we do not condition on the total number of books $N$. Thus the conditional distrbution is the same as the unconditional distribution which is $Pois(\lambda p)$.

47. (a) $N_t \sim Pois(\lambda t)$ so the mean and variance are $\lambda t$.

(b) Let $A_i$ be the number of passengers on bus $i$. Then $X_t = A_1 + \cdots + A_{N_t}$ and $E(X_t) = E(E(X_t|N_t)) = E(N_t np) = np\lambda t$ and $Var(X_t) = E(Var(X_t|N_t)) + Var(E(X_t|N_t)) = E(N_t np(1-p)) + Var(N_t np) = \lambda t np(1-p) + n^2p^2\lambda t$.

(c) Let $A$ be the number of full buses and $B$ the number of not full buses. The probability that an individual bus is full is $p^n$. Then by the chicken egg story, the number of full buses is independent of the number of not full buses, so $P(N_t = a+b, A = a, B = b) = P(A = a, B = b) = P(A = a)P(B = b)$ where $A \sim Pois(\lambda t p^n)$ and $B \sim Pois(\lambda t(1-p^n))$.

48. (a) The $A_j$ are not independent, since knowing that some people finished faster than Paul makes it more likely that Paul took a long time to finish, so other people are also likely to have finished faster than Paul.

Given Paul's time $T$, the $A_j$ are independent since $P(A_j = 1|T) = F(T)$.

(b) $E(N) = E(\sum_j A_j) = \sum_j E(A_j) = n/2$ since the (unconditional) outcomes of any two people are equally likely to have any ordering.

(c) Given $T = t$, the conditional distribution of $N$ is $Bin(n, F(t))$.

(d) $Var(N) = E(Var(N|T)) + Var(E(N|T)) = E(nF(T)(1-F(T))) + Var(nF(T))$. By universality of the uniform, $F(T) \sim Unif(0,1)$ so $E(nF(T)(1-F(T))) = nE(U - U^2) = nE(U) - nE(U^2) = n/2 - n(Var(U) + E(U)^2) = n/2 - n/3 = n/6$ and $Var(nF(T)) = Var(nU) = n^2/12$. Thus $Var(N) = n/6 + n^2/12$.

49. (a) Let $A$ be the number of emails in the first two hours and $B$ the number of emails in the next six. Then since they are disjoint increments in a Poisson process, $A \sim Pois(1/4(8\lambda))$ and $B \sim Pois(3/4(8\lambda))$ are independent by the chicken-egg story. Furthermore, $A + B = C$ where $C \sim Pois(8\lambda)$. Then given $C$, $A$ and $B$ are conditionally dependent with conditional distributions $A|C \sim Bin(C, 1/4)$ and $B|C \sim Bin(C, 3/4)$.

(b) By the chicken-egg story, the distributions of spam and non-spam emails are independent $Pois(8\lambda(1-p))$ and $Pois(8\lambda p))$ respectively. Thus given $s$ spams, independence implies that the conditional distribution of non-spam is the same as the unconditional distribution which is $Pois(8\lambda p)$.

(c) Let $C \sim Pois(8\lambda)$ be the number of emails that arrive in 8 hours. Then the expected reading time is $T_1 + \cdots + T_C$ where $T_i$ is the amount of time to read the $i^{th}$ email. Then $E(T) = E(E(T|C)) = E(C\mu) = 8\lambda\mu$ and $Var(T) = E(Var(T|C)) + Var(E(T|C)) = E(C\sigma^2) + Var(C\mu) = 8\lambda\sigma^2 + 8\lambda\mu^2$.

50. (a) $E(N) = E(E(N|\lambda)) = E(\lambda) = 1$. $Var(N) = E(Var(N|\lambda)) + Var(E(N|\lambda)) = E(\lambda) + Var(\lambda) = 1 + 2 = 3$.

(b) Let the amount of claim $i$ be $A_i$ and the total amount be $A$. Then $E(A) = E(E(A|N)) = E(Ne^{\mu+\sigma^2/2}) = e^{\mu+\sigma^2/2}$ and $Var(A) = E(Var(A|N)) + Var(E(A|N)) = E(N(e^{\sigma^2} - 1)(e^{2\mu+\sigma^2})) + Var(Ne^{\mu+\sigma^2/2}) = (e^{\sigma^2} - 1)(e^{2\mu+\sigma^2}) + 3e^{2\mu+\sigma^2}$.

(c) $E(e^{tN}) = E(E(e^{tN}|\lambda)) = E(e^{\lambda(e^t-1)})$ since the $\sum_k e^{tk}e^{-\lambda}\lambda^k/k! = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}$. Then continuing, $= \int_0^\infty e^{x(e^t-1)}e^{-x}dx = \int_0^\infty e^{x(e^t-2)}dx = 1/(2-e^t) = (1/2)/(1-1/2e^t)$ which is the MGF of a $Geom(1/2)$ random variable.

(d) By Gamma Poisson conjugacy, $\lambda \sim Gamma(1,1)$ is the prior and $\lambda|N = n \sim Gamma(n+1,2)$.

51. (a) By independence and linearity, the expectation is equal to $\sum_i E(X_i|X_i > i)$. By the memoryless property of exponentials, this equals $\sum_i i + E(X_i) = 6 + 1/\lambda_1 + 1/\lambda_2 + 1/\lambda_3$.

(b) The straightforward approach would be to evaluate a triple integral, which I'm sure would yield $\lambda_1/(\lambda_1+\lambda_2+\lambda_3)$. The hint provides a clever alternative: $\min(X_1,X_2,X_3) = \min(X_1, \min(X_2,X_3))$. By properties of the exponential, $\min(X_2,X_3) \sim Expo(\lambda_2+\lambda_3)$ and then by the given result $P(X_1 = \min(X_1,X_2,X_3)) = P(X_1 \leq \min(X_2,X_3)) = \lambda_1/(\lambda_1+\lambda_2+\lambda_3)$.

(c) $\max(X_1,X_2,X_3)$ is waiting until the first event, then the second event, then the final event. If $\lambda_1,\lambda_2,\lambda_3$ are distinct, this requires some conditioning. However, given they're all equal, then the waiting time until the first event is $Expo(3)$, the waiting time until the second is $Expo(2)$, and the waiting time until the last is $Expo(1)$, so $\max(X_1,X_2,X_3) \sim Expo(1) + Expo(2) + Expo(3)$.

52. (a) $E(T) = 1/2E(T_1) + 1/2E(T_2) = 1/2(1/\lambda_1 + 1/\lambda_2)$. $Var(T) = E(Var(T|J)) + Var(E(T|J)) = E(1/\lambda_J^2) + Var(1/\lambda_J) = 1/2(1/\lambda_1^2 + 1/\lambda_2^2) + 1/2(1/\lambda_1^2 + 1/\lambda_2^2) - 1/4(1/\lambda_1 + 1/\lambda_2)^2 = (1/\lambda_1^2 + 1/\lambda_2^2) - (1/\lambda_1^2 + 2/\lambda_1\lambda_2 + 1/\lambda_2^2)/4$.

(b) $X = \min(T_1,T_2) \sim Expo(\lambda_1+\lambda_2)$. $E(X|X \geq 24) = 24 + E(X) = 24 + 1/(\lambda_1+\lambda_2)$.

53. (a) $E(Y) = E(E(Y|\mu)) = E(\mu) = 100$. $Var(Y) = E(Var(Y|\mu)) + Var(E(Y|\mu)) = 25 + Var(\mu) = 25 + 225 = 250$.

(b) $E(e^{tY}) = E(E(e^{tY}|\mu)) = E(e^{\mu t + t^2 5^2/2}) = e^{25t^2/2}E(e^{t\mu}) = e^{25t^2/2}e^{25t^2/2}e^{100t+225t^2/2} = e^{100t+250t^2/2}$ so $Y \sim N(100, 250)$.

(c) $Cov(\mu, Y) = E(\mu Y) - E(\mu)E(Y) = E(E(\mu Y|\mu)) - 100E(E(Y|\mu)) = E(\mu^2) - 100E(\mu) = 15^2 + 100^2 - 100^2 = 15^2$.

54. $2Cov(X_1,X_2) = 2E(X_1X_2) - 2E(X_1)E(X_2) = E(X_1X_2|I) + E(X_1X_2|F) - 0 = E(X_1^2) + \rho\sigma^2 = \sigma^2(\rho+1)$.

55. If we are missing $j$ out of 35 numbers, then we can draw up to $\min(5,j)$ of them. The probabilities are $\binom{35-j}{5-i}\binom{j}{i}/\binom{35}{5}$ of choosing $i$ new numbers, so $a_j = 1 + \sum_{i=0}^{\min(5,j)} \binom{35-j}{5-i}\binom{j}{i}/\binom{35}{5}a_{j-i}$ for $j > 0$.

56. (a) Let $G$ be the number of games until Vishy wins, so $G|p \sim FS(p)$. Unconditionally, $E(G) = E(E(G|p)) = E(1/p) = \int_0^1 1/\beta(a,b)1/pp^{a-1}(1-p)^{b-1}dp = \beta(a-1,b)/\beta(a,b) = (a+b-1)/(a-1)$ by pattern recognition.

(b) $1 + E(G) = E(FS(a/(a+b))) = (a+b)/a = 1/E(p)$ and we are comparing this to the value computed previously of $E(1/p)$. They are equal if and only if $p$ and $1/p$

are uncorrelated. However, $p$ and $1/p$ are negatively correlated, since $p$ increases if and only if $1/p$ decreases. Thus $Cov(p, 1/p) < 0 \implies 1 - E(p)E(1/p) < 0 \implies 1/E(p) < E(1/p)$.

(c) By beta-binomial conjugacy, the posterior (conditional) distribution of $p$ is $Beta(a+7, b+3)$.

57. (a) The posterior distribution of $p$ given $X_1, \ldots, X_n$ is $Beta(1 + X_1 + \cdots + X_n, 1 + (1 - X_1) + \cdots + (1 - X_n))$, which depends only on the sum, by beta-binomial conjugacy.

(b) $P(X_{n+1} = 1|X_1 + \cdots + X_n = k) = \int_0^1 P(X_{n+1} = 1|p, X_1 + \cdots + X_n = k)f(p|X_1 + \cdots + X_n = k)dp = \int_0^1 p(1/\beta(1+k, 1+n-k))p^k(1-p)^{n-k}dp = \beta(k+2, n-k+1)/\beta(k+1, n-k+1) = (k+1)/(n+2)$.

58. (a) $E(X) = E(E(X|p)) = E(np) = n/2$. $Var(X) = E(Var(X|p)) + Var(E(X|p)) = E(np(1-p)) + Var(np) = (n/2 - n/3) + n^2/12$.

(b) Let $I_j$ be the indicator of team $A$ winning game $j$. Then $P(I_{j+1} = 1|a) = E(I_{j+1}|a) = E(E(I_{j+1}|a, p)|a) = E(p|a)$. Given $a$, $p \sim Beta(1+a, 1+j-a)$ by beta-binomial conjugacy, so $E(p|a) = (1+a)/(j+2)$.

(c) By beta-binomial conjugacy with prior $p \sim Unif(0,1) \sim Beta(1,1)$, we have $P(X = x) = 1/(n+1)$ so $X \sim Dunif(0, \ldots, n)$.

(d) The probability is 1/99 (as it would have been for any $m$, not just 50). Part (c) applies, since given the first two shots, the prior going forward for the remaining 98 shots is $Beta(1,1) = Unif(0,1)$.

59. $E(X) = E(E(X|p)) = E(np) = na/(a+b)$. $Var(X) = E(Var(X|p)) + Var(E(X|p)) = E(np(1-p)) + Var(np) = na/(a+b) - n(a+1)a/((a+b+1)(a+b)) + n^2((a+1)a/((a+b+1)(a+b)) - a^2/(a+b)^2)$.

60. (a) By problem 58, $X_1 \sim Dunif(0, n_1)$. The posterior distribution of $p_1|(X_1 = k_1)$ is $Beta(k_1 + 1, n_1 - k_1 + 1)$.

(b) $E(X) = \sum_j E(X_j) = \sum_j E(E(X_j|p_j)) = \sum_j E(n_j p_j) = \sum_j n_j/2 = n/2$. $Var(X) = \sum_j Var(X_j) = \sum_j E(Var(X_j|p_j)) + Var(E(X_j|p_j)) = \sum_j E(n_j p_j(1-p_j)) + Var(n_j p_j) = \sum_j n_j(1/2 - 1/3) + n_j^2/12 = n/6 + s/12$.