1 Change of variables

1. $0 < e^{-X} < 1$ so the support of the PDF is $0 < x < 1$. Then $P(e^{-X} \le x) = P(X \ge -\log(x)) = 1 - P(X \le -\log(x)) \implies f(x) = -\frac{d}{dx}(1 - P(X \le -\log(x))) = -f_X(-\log(x))(-1/x) = f_X(-\log(X))/x = e^{-(-\log(x))}/x = 1$. We could also have done this by universality of the uniform since $e^{-X} = 1 - F_X(X)$ and $F_X(X) \sim Unif(0,1)$ and $1 - Unif(0,1) \sim Unif(0,1)$.

2. The support of the PDF is $0 < x < \infty$. $P(X^7 \le x) = P(X \le \sqrt[7]{x}) = F(\sqrt[7]{x})$ so the PDF is $f_X(\sqrt[7]{x})x^{-6/7}/7 = \lambda e^{-\lambda \sqrt[7]{x}} x^{-6/7}/7$.

3. The support of the PDF is $-\infty < z < \infty$. $P(Z^3 \le z) = P(Z \le \sqrt[3]{z}) = \Phi(\sqrt[3]{z})$ so the PDF is $\varphi(\sqrt[3]{z})z^{-2/3}/3 = \frac{1}{\sqrt{2\pi}} e^{-z^{2/3}/2} z^{-2/3}/3$.

4. The support of the PDF is $0 \le z < \infty$. $P(Z^4 \le z) = P(-\sqrt[4]{z} \le Z \le \sqrt[4]{z}) = \Phi(\sqrt[4]{z}) - \Phi(-\sqrt[4]{z}) = 1 - 2\Phi(-\sqrt[4]{z}) = 2\Phi(\sqrt[4]{z}) - 1$ so the PDF is $2\varphi(\sqrt[4]{z})z^{-3/4}/4 = \sqrt{2/\pi} e^{-z^{1/2}/2} z^{-3/4}/4$.

5. The support of the PDF is $0 \le z < \infty$. $P(|Z| \le z) = P(-z \le Z \le z) = 2\Phi(z) - 1$ so the PDF is $2\varphi(z) = \sqrt{2/\pi} e^{-z^2/2}$.

6. The supports are both $(0, 1)$. (Note that we could apply the change of variables formula directly since both functions are increasing on the support.) The PDF of the first is $P(U^2 \le u) = P(U \le \sqrt{u})$ so the PDF is $u^{-1/2}/2$.

   The PDF of the second is $P(\sqrt{U} \le u) = P(U \le u^2)$ so the PDF is $2u$.

7. The support of the PDF is $(0, 1)$. (Again, we could apply the change of variables formula directly.) $P(\sin(U) \le u) = P(U \le \sin^{-1}(u))$ so the PDF is $2/pix/\sqrt{1 - x^2}$.

8. (a) $P(X^2 = x) = P(X = \sqrt{x}) = 1/n$ if $x = i^2$ for $0 \le i \le n$. Thus $X^2 \sim Dunif(0, 1, 4, \ldots, n^2)$.

   (b) $P(X^2 = x) = P(X = \pm\sqrt{x}) = 2/n$ if $x = i^2$ for $1 \le i \le n$ and $1/n$ if $x = 0$ and 0 otherwise.

9. $Y = (b - a)X + a$ satisfies the requirements. $P(Y = b) = P((b - a)X + a = b) = P((b - a)X = b - a) = P(X = 1) = p$ and $P(Y = a) = P((b - a)X + a = a) = P((b - a)X = 0) = P(X = 0) = (1 - p)$.

10. $P(Y = 1) = \sum_{k \ge 0} e^{-\lambda}\lambda^{2k+1}/(2k+1)!$ and $P(Y = 0) = \sum_{k \ge 0} e^{-\lambda}\lambda^{2k}/(2k)!$. $P(Y = 1) + P(Y = 0) = 1$ since it is the total probability of the Poisson PMF while $P(Y = 0) - P(Y = 1) = \sum_{k \ge 0} e^{-\lambda}(-\lambda)^k/k! = e^{-2\lambda}$. Thus $P(Y = 1) = (1 - e^{-2\lambda})/2$.

11. If $v = 0$ and $T < 0$, then $1/T < 0$ so $V \le 0$ if and only if $T \le 0$ so $P(V \le 0) = P(T \le 0)$.

    If $v > 0$ and $T > 0$, then $1/T < v \iff T > 1/v$ so $P(V \le v) = P(T > 1/v) = P(T \le 0) + P(T > 1/v)$.

    Finally, if $v < 0$ and $T < 0$, then $1/T < v \iff 1 > Tv \iff 1/v < T$ so $P(V \le v) = P(T > 1/v) = P(T \le 0) - P(T \le 1/v)$.

12. (a) $f_V(v) = f_T(1/v)/v^2 = 1/\pi 1/(1+v^2) = f_T(v)$ so they have the same distribution.
    (b) X/Y is equivalent in distrbution to Y/X since X and Y are iid.

13. $P(T \leq t) = P(\log(X/Y) \leq t) = P(X \leq (e^t)Y) = \int_0^\infty \int_0^{e^t y} \lambda e^{-\lambda x} \lambda e^{\lambda y} dxdy = 1 - 1/(e^t + 1)$. The PDF is $e^t/(1+e^t)^2$.

14. (a) By 2d change of variables, $f_{T,W}(t,w) = f_{X,Y}((dt-bw)/(ad-bc), (-ct+aw)/(ad-bc)) * 1/|ad-bc|$.
    (b) Plugging in $a = b = c = 1, d = -1$ to the above yields the desired result.

15. (a) Knowing θ gives no information about R since X, Y are spherically symmetric and thus depend only on the radial distance and not on the angle.
    (b) By $2d$ change of variables, $f_{R,\theta}(r,\theta) = f_{X,Y}(r\cos\theta, r\sin\theta)r = r/\pi$. (Note that when we integrate, we are doing a normal cartesian integral rather than a polar integral, even though f is considered in polar coordinates.)
    (c) $f_{R,\theta}(r,\theta) = 1/2\pi e^{-r^2\cos^2\theta/2} e^{-r^2\sin^2\theta/2} r = 1/2\pi e^{-r^2/2} r$.

16. Using the result of exercise 14(b), $f_{T,W}(t,w) = 1/2 f_{X,Y}((t+w)/2, (t-w)/2) = 1/4\pi e^{-(t+w)^2/2-(t-w)^2/2} = 1/4\pi e^{-1/2((t^2+2tw+w^2)+t^2-2tw+w^2)/2} = 1/4\pi e^{-t^2-w^2}$. Thus T, W are independent $N(0,2)$ random variables since the joint PDF factors as the product of the corresponding marginals.

17. $f_{R^2,\theta}(r,\theta) = f_{X,Y}(\sqrt{r}\cos\theta, \sqrt{r}\sin\theta)/2 = 1/4\pi e^{-r\cos^2/2 - r\sin^2/2} = 1/4\pi e^{-r/2} = (1/2\pi)1/2e^{-1/2r}$ so $R^2 \sim Expo(1/2)$ and $\theta \sim Unif(0, 2\pi)$.

18. (a) Let $W = X$. By the 2d change of variables, $f_{T,W}(t,w) = f_{X,Y}(w, w/t)w/t^2 = (wf_X(w))(f_Y(w/t)/t^2)$.
    (b) $f_T(t) = \int_0^\infty f_{T,W}(t,w)dw = (wf_X(w))(f_Y(w/t)/t^2)dw$.

19. (a) The inverse transform is $X = TW/(W+1), Y = T/(W+1)$ and the determinant of the Jacobian is $T/(W+1)^2$, so the change of variables formula gives $f_{T,W}(t,w) = f_{X,Y}(tw/(w+1), t/(w+1))t/(w+1)^2 = \lambda e^{-\lambda tw/(w+1)} \lambda e^{-\lambda t/(w+1)} t/(w+1)^2 = t\lambda^2 e^{-\lambda t}(1+w)^{-2}$. This factors so T and W are independent.
    (b) Manipulating the joint PDF yields: $(\lambda t)^2/te^{-\lambda t}(1+w)^{-2}$ so the marginal distribution of T is $Gamma(2,\lambda)$ while the marginal distribution of W is $f_W(w) = 1/(1+w)^2$.

2 Convolutions

20. $P(U + X \leq t) = \int_0^1 \int_0^{t-u} e^{-x} dxdu = \int_0^1 1 - e^{u-t} du = 1 - e^{1-t} + e^{-t}$.

21. When $l > 0$, $P(L \leq l) = P(X \leq Y + l) = \int_0^\infty \int_0^{y+l} e^{-x} e^{-y} dxdy = \int_0^\infty (1 - e^{-l}e^{-y})e^{-y} dy = 1 - e^{-l}/2$. When $l < 0$, $P(L \leq l = P(X \leq Y + l)) = \int_{-l}^\infty \int_0^{y+l} e^{-x}e^{-y} dxdy = \int_{-l}^\infty (1 - e^{-y-l})e^{-y} dy = e^l - 1/2e^l = 1/2e^l$. Taking derivatives, we obtain $f(t) = e^{-l}/2 = e^{-|l|}/2$ when $l > 0$ and $f(t) = e^l/2 = e^{-|l|}/2$ when $l < 0$.

22. $P(T \leq t) = \int_{-\infty}^{\infty} \int_{-\infty}^{t-x} \varphi(x)\varphi(y)dydx = \int_{-\infty}^{\infty} \varphi(x)\Phi(t-x)dx$ so by differentiating under the integral sign, the PDF is $\int_{-\infty}^{\infty} \varphi(x)\varphi(t-x)dx$....whose calculation would yield a $N(\mu_1+\mu_2, 2\sigma^2)$ PDF.

23. (a) $P(XY \leq t) = P(X \leq t/Y) = \int_0^\infty P(X \leq ty|Y=y)f(y)dy = \int_0^\infty F_X(ty)f(y)dy$. Taking a derivative yields $f_T(t) = \int_0^\infty y f_X(ty)f(y)dy$.

    (b) $\log(T) = \log(X) + \log(Y)$ so $P(T \leq t) = P(\log(T) \leq \log(t)) = P(\log(X) + \log(Y) \leq \log(t))$. Let $A = \log(X), B = \log(Y)$ so we continue with $= P(A+B \leq \log(t)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\log(t)-a} f_A(a)f_B(b)dbda = \int_{-\infty}^{\infty} f_A(a)F_B(\log(t)-a)da$. Taking a t derivative yields $\int_{-\infty}^{\infty} f_a(a)f_B(\log(t)-a)/tda$.

    $P(A \leq a) = P(X \leq e^a) = F_X(e^a)$ so the PDF of A is $e^a f_X(e^a)$ and analogously for Y. Thus we obtain $\int_{-\infty}^{\infty} e^a f_X(e^a) e^{\log(t)-a} f_Y(e^{\log(t)-a})/tda = \int_{-\infty}^{\infty} f_X(e^a)f_Y(e^{\log(t)-a})da$. Converting back to an integral with respect to $x = e^a \implies dx = e^a da \implies da = dx/x$ yields $\int_0^\infty f_X(x)f_Y(t/x)/xdx$. (The other integral can be reobtained by setting $y = t/x$.)

24. $P(T=t) = \sum_k P(X+Y=t|X=k)P(X=k) = 1/(n+1)\sum_k 1/(n+1)$ so the only question is how many terms the sum contains. For $0 \leq t \leq n$, there are $t+1$ possible ways. For $n+1 \leq t \leq 2n$, there are $2n+1-t$ ways. Thus $P(T=t) = (t+1)/(n+1)^2$ if $0 \leq t \leq n$ and $(2n+1-t)/(n+1)^2$ if $n+1 \leq t \leq 2n$.

25. (a) $E(X-Y) = E(X) - E(Y) = 0$ and $Var(X-Y) = Var(X) + Var(-Y) = 1/6$.

    (b) $-W$ has the same distribution as W since $X - Y \sim Y - X$ are both a difference of iid $Unif(0,1)$ random variables.

    (c) $P(W \leq w) = P(X - Y \leq w) = P(X \leq Y + w)$. For $w > 0$, this is equal to $\int_0^1 \int_0^{\min(1,y+w)} dxdy = \int_0^1 \min(1, y+w)dy = 1-(1-w)^2/2$. For $w < 0$, this is equal to $\int_{-w}^1 \int_0^{y+w} dxdy = \int_{-w}^1 y + wdy = 1/2 - w^2/2 + w + w^2 = 1/2(w^2+2w+1) = 1/2(w+1)^2$. The PDF is then $1 - w = 1 - |w|$ if $w > 0$ and $1 + w = 1 - |w|$ if $w < 0$ so the overall pdf is $1 - |w|$.

    (d) $E(W) = \int_{-1}^1 w(1-|w|)dw = \int_{-1}^0 w(1+w)dw + \int_0^1 w(1-w)dw = -1/6 + 1/6 = 0$. $Var(W) = \int_{-1}^1 w^2(1-|w|)dw = \int_{-1}^0 w^2(1+w)dw + \int_0^1 w^2(1-w)dw = 1/12 + 1/12 = 1/6$. The distribution is symmetric since the PDF is even.

    (e) W is a shifted version of the triangle distribution. Let $Z = 1 - Y$. Then $Z \sim Unif(0,1)$ is independent of X, so $1 + W = X + (1-Y) = X + Z = T$.

26. The area under the line $x + y = t$ and in the unit square is the probability that $X + Y \leq t$. For $t \leq 1$, this is an isosceles right triangle with side lengths t so the area is $t^2/2$. For $t \geq 1$, this is the complement of an isosceles right triangle with side lengths $2 - t$ so the area is $1 - (2-t)^2/2$.

27. For $0 \leq w \leq 1$, then $P(X+Y+Z \leq w) = P(T + Z \leq w) = \int_0^w \int_0^{w-x} tdtdx = \int_0^w (w - x)^2/2dx = w^3/6$. For $1 \leq w \leq 2$, then $P(T+Z \leq w) = \int_0^w \int_0^{\min(1,w-t)} f_T(t)dxdt = \int_0^w f_T(t)\min(1, w-t)dt = \int_0^1 t\min(1, w-t)dt + \int_1^w (2-t)\min(1,w-t)dt = \int_0^{w-1} t(w-$

$t)dt + \int_{w-1}^{1} tdt + \int_{1}^{w} (2-t)(w-t)dt = (w-1)^2(w+2)/6 + 1/2(1-(w-1)^2) + 2w(w-1) - (w+2)(w^2-1)/2 + (w^3-1)/3 = 1/2(w^2-2w+2)$. Finally, for $2 \le w \le 3$, then $P(T+Z \le w)1 - P(T+Z > w) = \int_{w-1}^{2}\int_{t-w}^{1}(2-t)dxdt = \int_{w-1}^{2}(1-t+w)(2-t)dt = 2(2-w+1) - 3/2(4-(w-1)^2) + 1/3(8-(w-1)^3) + 2w(2-w+1) - w/2(4-(w-1)^2) = 1/6(w-3)^2(w+3)$. (Probably made quite a few computational errors here, but the idea is straightforward enough...)

3 Beta and Gamma

28. (a) $P(1-B \le x) = P(B \ge 1-x) = 1 - F(1-x)$. Thus the pdf is $f(1-x)$ where $f(x) = 1/\beta(a,b)x^{a-1}(1-x)^{b-1}$ is the $Beta(a,b)$ PDF. Thus the pdf is $f(1-x) = 1/\beta(a,b)(1-x)^{a-1}x^{b-1}$ which is the $Beta(b,a)$ PDF.

    (b) Thinking of a and b as the number of successes and failures in a long sequence of Bernoulli trials and then of $Beta(a,b)$ as the distribution of the probability of success for each Bernoulli trial, then $1 - Beta(a,b)$ should represent the distribution of the probability of failure for the same Bernoulli trials, so we should switch a and b.

29. (a) $f_T(t) = \int_0^\infty f_X(t-y)f_y(y)dy = \int_0^\infty (1/\Gamma(a)(\lambda(t-y))^a e^{-\lambda(t-y)}/(t-y))(1/\Gamma(b)(\lambda y)^b e^{-\lambda y}/y)dy = (\lambda t)^{a+b}e^{-\lambda}/(t\Gamma(a)\Gamma(b))\int_0^\infty((1-z))^{a-1}(z)^{b-1}dz$. The inner integral is now the $Beta(a,b)$ integral, so its value is $\Gamma(a)\Gamma(b)/(\Gamma(a+b))$ hence the result is $(\lambda t)^{a+b}e^{-\lambda}/(t\Gamma(a+b))$ which is the $Gamma(a+b,\lambda)$ integral.

    (b) The MGFs of X and Y are $(\lambda/(\lambda-t))^a$ and $(\lambda/(\lambda-t))^b$ so the product is $(\lambda/(\lambda-t))^{a+b}$ which is the $Gamma(a+b,\lambda)$ MGF.

    (c) One way to realize $Gamma(a,\lambda)$ is as the sum of a iid $Expo(\lambda)$ rvs. Thus, $X+Y$ can be realized as the sum of $a+b$ iid $Expo(\lambda)$ rvs so it has the $\Gamma(a+b,\lambda)$ distribution. (Thus this is the time until $a+b$ occurrences.)

30. $E(B^k) = \int_0^1 x^k 1/\beta(a,b)x^{a-1}(1-x)^{b-1}dx = \int_0^1 1/\beta(a,b)x^{a+k-1}(1-x)^{b-1}dx = \beta(a+k,b)/\beta(a,b) = \Gamma(a+b)/\Gamma(a)\Gamma(a+k)/\Gamma(a+k+b) = (a+k-1)\cdots(a)/(a+k+b-1)\cdots(a+b)$.

    In particular, for $k=2$, this is $(a+1)(a)/(a+b+1)(a+b)$, so the variance is $(a+1)(a)/(a+b+1)(a+b) - a^2/(a+b)^2 = ab/(a+b+1)(a+b)^2$.

31. The inverse transform (as in 19) is $X = TW/(W+1), Y = T/(W+1)$. Thus by 2d change of variables, $f_{T,W}(t,w) = f_{X,Y}(tw/(w+1), t/(w+1))t/(w+1)^2 = 1/\Gamma(a)(\lambda tw/(w+1))^a e^{-\lambda tw/(w+1)}/(tw/(w+1))1/\Gamma(b)(\lambda t/(w+1))^b e^{-\lambda t/(w+1)}/(t/(w+1)) * t/(w+1)^2 = 1/\Gamma(a)\Gamma(b)(\lambda t)^{a+b}e^{-\lambda t}/t * w^{a-1}/(w+1)^{a+b}$. Since the pdf factors, they are independent (similarly to problem 19 in the case $a=b=1$).

    Alternatively, $X/Y = (X/(X+Y))/(1-X/(X+Y)) = W/(1-W)$ in the bank-post office story where W is the ratio of time spent waiting for X to the total amount of time spent waiting. Since W is independent of $X+Y$, then $W/(1-W)$ is also independent of $X+Y$.

32. Simplifying, we find $mV/(n+mV) = 1/(1+Y/X) = X/(X+Y)$. Since X, Y are both positive, then $1+Y/X > 1$ so $mV/(n+mV)$ has support $(0,1)$, and we could try to find the answer by integrating against the PDF of Y/X which we found in the previous problem. since $P(mV/(n+mV) \leq v) = P(Y/X \geq 1/v - 1) = 1 - F_{Y/X}(1/v-1)$ so the derivative with respect to v yields $f_{Y/X}(1/v-1)/v^2$.

    However, an easier approach is to recognize $X/(X+Y)$ as the ratio from the bank-post office story. In that story, $W \sim Beta(m/2, n/2)$.

33. This is the computation we already did with Gamma-Poisson conjugacy, so since we observed 2 customers in two hours, the updated posterior PDF is $Gamma(3,5)$.

34. (a) Let $X \sim Unif(0,1)$ and $Y \sim Unif(0,2)$. Then $E(X/(X+Y)) = \int_0^2 \int_0^1 x/2(x+y)dxdy$. Using some manipulation and integration by parts, the result is $1/2 - 3\log(3)/4 + \log(2) \approx .369188$ while $E(X)/(E(X)+E(Y)) = .5/(.5+1) = 1/3$. (Should have probably used a discrete example...)

    (b) $Cov(X/(X+Y), X+Y) = Cov(X/(X+Y), X) + Cov(X/(X+Y), Y) = Cov(X/(X+Y), X) + Cov(1-Y/(X+Y), Y) = Cov(X/(X+Y), X) - Cov(Y/(X+Y), Y) = 0$ since X, Y are iid. Thus $E(X) - E(X/(X+Y))E(X+Y) = 0 \implies E(X)/E(X+Y) = E(X/(X+Y))$ if X and Y are iid.

    (c) By the bank-post office story, $X/(X+Y)$ and $X+Y$ are independent, so $E(X^c) = E((X/(X+Y))^c(X+Y)^c) = E((X/(X+Y))^c)E((X+Y)^c) \implies E(X^c/(X+Y)^c) = E(X^c)/E((X+Y)^c)$.

35. $X \sim Gamma(1, \lambda)$. Since $\gamma > 0$, then $1/\gamma > -1$ so the previous result yields $E(T) = 1/\lambda^{1/\gamma}\Gamma(1+1/\gamma)/\Gamma(1)$. For the variance, $E(T^2) = 1/\lambda^{2/\gamma}\Gamma(1+2/\gamma)/\Gamma(1)$, so $Var(T) = 1/\lambda^{2/\gamma}\Gamma(1+2/\gamma)/\Gamma(1) - (1/\lambda^{1/\gamma}\Gamma(1+1/\gamma)/\Gamma(1))^2 = 1/\lambda^{2/\gamma}\Gamma(1+2/\gamma) - 1/\lambda^{2/\gamma}\Gamma(1+1/\gamma)^2$

36. (a) If $\lambda_1 = \lambda_2$, then by the bank-post office story, $T = T_1 + T_2$ and $W = T_1/(T_1+T_2)$ are independent. Then T and $W/(1-W) = T_1/T_2$ are independent.

    (b) $T_1 < T_2$ if and only if $T_1/T_2 = W/(1-W) < 1$. If $\lambda_1 = \lambda_2$, then by the bank-post office story, $W \sim Beta(1,1) = Unif(0,1)$ since $Expo(\lambda) = Gamma(1, \lambda)$, so $P(T_1 < T_2) = 1/2$. Otherwise, $P(T_1/T_2 \leq t) = P(T_1 \leq tT_2) = \int_0^\infty \int_0^{tx} \lambda_1 e^{-\lambda_1 y}\lambda_2 e^{-\lambda_2 x}dydx = \int_0^\infty (1-e^{-\lambda_1 tx})\lambda_2 e^{-\lambda_2 x}dx = \lambda_1 t/(\lambda_1 t + \lambda_2)$ so $P(T_1 < T_2) = \lambda_1/(\lambda_1+\lambda_2)$.

    (c) Alice first waits for $\min(Expo(\lambda_1), Expo(\lambda_2))$ time. Now let X be the time that it takes Alice to be served. Then using LOTP, $X = \lambda_1/(\lambda_1+\lambda_2)T_1 + \lambda_2/(\lambda_1+\lambda_2)T_2$ so the expected total amount of time that Alice spends is $E(\min(Expo(\lambda_1), Expo(\lambda_2)) + \lambda_1/(\lambda_1+\lambda_2)T_1 + \lambda_2/(\lambda_1+\lambda_2)T_2)$. Since $\min(Expo(\lambda_1), Expo(\lambda_2)) \sim Expo(\lambda_1+\lambda_2)$, the total expected waiting time is $3/(\lambda_1+\lambda_2)$.

37. Suppose we are waiting for arrivals in a period of length t with arrival rate λ. Then the waiting time until j arrivals is a sum of j iid $Expo(1, \lambda)$ random variables which is distributed as $Gamma(j, \lambda)$. Then the event that there are more than or equal to j arrivals in the length t time period is equal to the event that the time before j arrivals is less than or equal to t, so $P(X \geq j) = P(Y \leq t)$.

38. Let T be Fred's time at the park and N be the total number of visitors during that time. Then $T \sim Expo(\lambda_2)$ and $N|T = t \sim Pois(\lambda t)$. To find the unconditional distribution of N, we need to integrate out t. $P(N = n) = \int_0^\infty P(N = n|T = t)f_T(t)dt = \int_0^\infty e^{-\lambda t}(\lambda t)^n/n!\lambda_2 e^{-\lambda_2 t}dt = \lambda_2\lambda^n/n!\int_0^\infty e^{-(\lambda+\lambda_2)t}t^n dt = \lambda_2\lambda^n/(\lambda + \lambda_2)^{n+1}/n!\int_0^\infty e^{-(\lambda+\lambda_2)t}((\lambda + \lambda_2)t)^{n+1}/tdt = \Gamma(n+1)\lambda_2\lambda^n/n!(\lambda + \lambda_2)^{n+1} = \lambda_2\lambda^n/(\lambda + \lambda_2)^{n+1}$. Thus $N \sim Geom(\lambda_2/(\lambda + \lambda_2))$.

39. (a) From a previous problem, $E(B^k) = (a + k - 1)\cdots a/(a + b + k - 1)\cdots(a + b)$. Then $E(p^2(1-2p+p^2)) = E(p^2) - 2E(p^3) + E(p^4) = (a+1)a/(a+b+1)(a+b) - 2(a+2)(a+1)a/(a+b+2)(a+b+1)(a+b) + (a+3)(a+2)(a+1)a/(a+b+3)(a+b+2)(a+b+1)(a+b) = a(a+1)b(b+1)/(a+b)(a+b+1)(a+b+2)(a+b+3)$.

    (b) The posterior distribution does not depend on the order of outcomes since we are assuming that p does not change over time. Thus we are in the situation of the beta-binomial conjugacy so the updated distribution depends only on the fact that A won 6 out of 10 games.

    (c) The posterior distribution for p is $Beta(7, 5)$ by the story of beta-binomial conjugacy.

    (d) Given p, the outcomes of the games are independent, so the first and second game are uncorrelated conditional on p.

    If we condition only on the historical data, then A winning the first game is positively correlated with A winning the second game. $P(A_1A_2) = \int_0^1 P(A_1A_2|p)f(p)dp = \int_0^1 1/\beta(7,5)p^8(1-p)^4dp = \beta(9,5)/\beta(7,5)$. Analogously, $P(A_1) = \beta(8,5)/\beta(7,5)$ so the covariance between winning games 1 and 2 is $\beta(9,5)/\beta(7,5) - \beta(8,5)^2/\beta(7,5)^2 = 8 * 7/13 * 12 - (7/12)^2 = 7/12(8/13 - 7/12) = 7/12 * (5/(12 * 13)) > 0$.

    (e) Let T be the event of a tie after the first four games, which requires that A wins 2 of the first four. This $P(T) = \int_0^1 P(T|p)f(p)dp = \int_0^1 \binom{4}{2}p^2(1-p)^2 1/\beta(7,5)p^6(1-p)^4dp = 6/\beta(7,5)\int_0^1 p^8(1-p)^6dp = 6\beta(9,7)/\beta(7,5) = 4/13$.

40. If all trials succeed, then $p \sim Beta(n + 1, 1)$. $P(p \geq r) = \int_r^1 1/\beta(n+1,1)p^n dp = \int_r^1 (n+1)p^n dp = 1 - r^{n+1}$.

4 Order statistics

41. An equivalent way to view the Binomial distribution is that we are tossing n balls uniformly randomly into $(0, 1)$ and want to find the probability that at least j of them are below p. Equivalently, we want to find the probability that the j^{th} order statistic is at most p. Since the distribution of the j^{th} order statistic is exactly $Beta(j, n - j + 1)$, the desired equality follows.

42. These three events are disjoint (since they each correspond to a different random variable being the minimum) and exactly one must occur since there must be a minimum between three numbers, so the probability is equal to 1.

43. The LHS integral is exactly $P(B \leq x)$ for $B \sim Beta(j, n-j+1)$ and the RHS is exactly $P(X \geq j)$ for $X \sim Bin(n,p)$.

44. By universality of the uniform, we can view each X_i as $F^{-1}(U_i)$ with inverse transformation $U_i = F(X_i)$. In particular, $F^{-1}(U_{(j)}) = X_{(j)}$ since F is strictly increasing so we can use the change of variables formula $f_{X_{(j)}}(x) = f_{Beta(j,n-j+1)}(F(x))f(x)$. Note that the RHS uses the known CDF and PDF of the X_i's rather than the unknown CDF/PDF of the order statistic.

45. (a) $P(M \leq m) = P(X \leq m)^2$. Then $P(X \leq m) = \int_0^m \lambda e^{-\lambda x}dx = 1 - e^{-\lambda m}$ so $P(M \leq m) = (1-e^{-\lambda m})^2$. On the other hand, $P(X + Y/2 \leq m) = \int_0^m \int_0^{m-x} \lambda e^{-\lambda x}(2\lambda)e^{-2\lambda y}dydx = \int_0^m \lambda e^{-\lambda x}(1-e^{-2\lambda(m-x)})dx = (1-e^{-\lambda m}) - (e^{-\lambda m} - e^{-\lambda 2m}) = (1-e^{-\lambda m})^2$.

    (b) We view the max as first waiting for one to happen, which is distributed as $\min(X,Y) \sim Expo(2\lambda)$. Then by the memoryless property, waiting for the other to happen takes another $Expo(\lambda)$ amount of time. Thus since $Y/2 \sim Expo(2\lambda)$, these two random variables have the same distribution.

46. (a) The continuous formula does not account for the possibility that $X = Y$ which can have nonzero probability in the discrete setting. $P(M = m) = P(X \leq m)P(Y = m) + P(X = m)P(Y \leq m) - P(X = m)P(Y = m) = 2F(m)P(X = m) - P(X = m)^2$.

    (b) There are three possibilities: $P(M = 1, L = 1) = 1/4, P(M = 1, L = 0) = 1/2, P(M = 0, L = 0) = 1/4$. The marginal PMFs are $P(M = 1) = 3/4$ and $P(L = 1) = 1/4$.

47. For $a \leq b$, $P(M_n \leq a, M_{n+1} \leq b) = P(M_{n+1} \leq b | M_n \leq a)P(M_n \leq a) = P(X_{n+1} \leq b)P(M_n \leq a) = F(b)F(a)^n$. For $a \geq b$, then $P(M_n \leq a, M_{n+1} \leq b) = P(M_n \leq a | M_{n+1} \leq b)P(M_{n+1} \leq b) = F(b)^{n+1}$.

48. If $a \geq b$ then $P(X_{(i)} \leq a, X_{(j)} \leq b) = P(X_{(i)} \leq a | X_{(j)} \leq b)P(X_{(j)} \leq b) = P(X_{(j)} \leq b)$. Otherwise, we repeat the argument used in the chapter: $f_{i,j}(a,b)dadb = nf(a)da(n-1)f(b)db\binom{n-2}{i-1,j-i-1,n-j}F(a)^{i-1}(F(b)-F(a))^{j-i-1}(1-F(b))^{n-j}$ and then dropping the differentials yields the desired PDF: $f_{i,j}(a,b) = nf(a)(n-1)f(b)\binom{n-2}{i-1,j-i-1,n-j}F(a)^{i-1}(F(b)-F(a))^{j-i-1}(1-F(b))^{n-j}$

49. (a) $L = \min(X,Y)$ where X, Y are the respective birth times. Note that $E(L+M) = E(X+Y) = 0$ where $M = \max(X,Y)$. Furthermore, $E(M-L) = E(|X-Y|) = \sqrt{2^7}E(|N(0,1)|)$ by properties of normal distributions. Then $E(|Z|) = \int_{-\infty}^0 -ze^{-z^2/2}/\sqrt{2\pi}dz + \int_0^\infty ze^{-z^2/2}/\sqrt{2\pi}dz = \sqrt{2/\pi}$ so $E(M-L) = \sqrt{2^8/\pi}$. Then solving the system of equations yields $E(L) = -1/2\sqrt{2^8/\pi} = -8/\sqrt{\pi}$.

    (b) $Var(T) = E(T^2) - 64/pi$. $E(T^2) = \int_{-\infty}^\infty \int_{-\infty}^\infty \min(x,y)^2\varphi(x)\varphi(y)dxdy$. To solve this, we can first subdivide into two integrals, one where $\min(x,y) = x$ and one where $\min(x,y) = y$. Then it's a straightforward integration problem...apparently the answer is 64.

50. (a) There are $n+1$ positions where Y_{new} can go relative to the order statistics of the $Y_1, \dots, Y_n$ and it is equally likely to land in any one of them. Thus the probability that $Y_{new} \in [Y_{(j)}, Y_{(k)}]$ is $(k-j)/n+1$.

    (b) We can take any $j < k$ which are at least 95 apart.

51. The median is the $m = (n+1)/2$ order statistic. The median of the sample median is $x_{(m)}$ such that $P(X_{(m)} \leq x_{(m)}) = P(X_{(m)} \geq x_{(m)}) = .5$. The median of the X_i's is x_m such that $P(X_i \leq x_m) = P(X_i \geq x_m) = .5$. Then $P(X_{(m)} \leq x_m)$ is equal to the probability that at least half the X_i are at most x_m, which has probability $\sum_{k=m}^{n} \binom{n}{k}(1/2)^k(1/2)^{n-k} = 1/2^n \sum_{k=m} \binom{n}{k} = 2^{n-1}/2^n = 1/2$. (This is because for n odd, $\sum_{k=m}^{n} \binom{n}{k} = \sum_{k=m}^{n} \binom{n}{n-k} = \sum_{k=0}^{m-1} \binom{n}{k}$.)

5 Mixed practice

52. (a) $P(X_j \leq x) = P(-\log(U_j) \leq x) = P(\log(U_j) \geq -x) = P(U_j \geq e^{-x}) = 1 - P(U_j \leq e^{-x}) = 1 - e^{-x}$. This is the CDF of the *Expo*(1) distribution.

    (b) $U = U_1 \cdots U_n$ and $L = -\log(U_1) - \cdots - \log(U_n)$. For $x \in (0,1)$, $P(U \leq x) = P(\log(U) \leq \log(x)) = P(-\log(U) \geq -\log(x)) = P(X_1 + \cdots + X_n \geq -\log(x))$. Since each $X_i \sim Expo(1)$, then $X_1 + \cdots + X_n \sim Gamma(n,1)$ so U has CDF $P(Gamma(n,1) \geq -\log(x)) = 1 - F_\Gamma(-\log(x))$ and PDF $f_\Gamma(-\log(x))/x$.

53. (a) Let I_j be the probability that starting at position j is CATCAT. Then $E(I_j) = (p_C p_A p_T)^2$ so $E(\sum_j I_j) = 110(p_C p_A p_T)^2$.

    (b) We can ignore T and G so $p_1/(p_1 + p_2)$.

    (c) By the beta-binomial conjugacy story, the posterior probability of C is $p_2 \sim Beta(2,3)$. The expected value is 2/5.

54. (a) X is negative binomial so can be considered as a sum of r iid geometric random variables. Then $M(t) = E(e^{tX}) = \prod_i E(e^{tX_i}) = (\frac{p}{1-(1-p)e^t})^r$. Thus $M(tp/(1-p)) = (\frac{p}{1-(1-p)e^{tp/(1-p)}})^r$. Computing $\lim_{p\to 0} M(tp/(1-p)) = (\lim_{p\to 0} \frac{p}{1-(1-p)e^{tp/(1-p)}})^r = (\lim_{p\to 0} \frac{1}{e^{tp/(1-p)}(1-t/(1-p))})^r = 1/(1-t)^r$.
    This is the MGF of the sum of r iid *Expo*(1) random variables or equivalently the MGF of the *Gamma*$(r,1)$ distribution.

    (b) We can view the *Gamma*$(r,1)$ distribution as the continuous limit of the Negative binomial distribution when the success probability goes to 0 (in a controlled way accounted for by the scaling $p/1-p$.