1 PDFs and CDFs

1. (a) $P(1 < X < 3) = \int_1^3 xe^{-x^2/2}dx = -e^{-x^2/2}|_1^3 = e^{-1/2} - e^{-9/2}$
   (b) $F(x) = \int_0^x xe^{-x^2/2}dx = -e^{-x^2/2}|_0^x = 1 - e^{-x^2/2}$. Then the inverse of $F^{-1}(x)$ is $u = F(x) = 1 - e^{-x^2/2}$ so $x = F^{-1}(u) = \sqrt{-2\log(1-u)}$. Then $q_j = F^{-1}(j/4)$.

2. (a) $f = 2$ on $[0, 1/2]$.
   (b) If $f(x) > 1$ on $[a, b]$, then $\int_a^b f(x)dx > \int_a^b 1dx = (b - a)$. Thus, $\int_{-\infty}^{\infty} f(x)dx > (b - a)$ if $f(x) > 1$ on (a, b). Thus the interval must have length smaller than 1, otherwise f will not be a valid PDF.

3. (a) $g(x)$ is nonnegative, so we only need to check its integral: $\int_{-\infty}^{\infty} 2F(x)f(x)dx = F(x)^2|_{-\infty}^{\infty} = 1 - 0 = 1$.
   (b) $h(x)$ is nonnegative, so we only need to check its integral: $\int_{-\infty}^{\infty} \frac{1}{2}f(-x)+\frac{1}{2}f(x)dx = \frac{1}{2}\int_{-\infty}^{\infty} f(-x)dx + \frac{1}{2}\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$.

4. (a) $P(X \le x|X > c) = P(c < X \le x)/P(c < X) = (F(x) - F(c))/(1 - F(c))$.
   (b) $f(X|X > c) = f(x)/(1 - F(c))$.
   (c) Nonnegativity is direct from the fact that f is a PDF and $F \le 1$ since it is a CDF. To see that it integrates to 1, we observe that as $x \to \infty$, we have $(1 - F(c))/(1 - F(c)) - (F(c) - F(c))/(1 - F(c)) = 1 - 0 = 1$.

5. (a) $A = \pi R^2$ so $E(A) = \pi E(R^2) = \pi(Var(R) + E(R)^2) = \pi(\frac{1}{12} + \frac{1}{4}) = \frac{\pi}{3}$. $Var(A) = \pi^2 Var(R^2) = \pi^2(E(R^4) - E(R^2)^2) = \int \pi^2 x^4 dx - \frac{\pi^2}{9} = \frac{4\pi^2}{45}$.
   (b) $P(A \le a) = P(\pi R^2 \le a) = P(R^2 \le \frac{a}{\pi}) = P(R \le \sqrt{\frac{a}{\pi}}) = \sqrt{\frac{a}{\pi}}$. Then $f(a) = \frac{1}{2\sqrt{a\pi}}$.

6. (a) The mean is $1/2$, so we want the probabilities $P(1/2 - i/\sqrt{12} < U < 1/2 + i/\sqrt{12}) = i/\sqrt{3}$ for $i = 1, 2, 3$.
   (b) The mean is $1 = 1/1$ and the variance is $1 = 1/1^2$, so we want the probabilities $P(0 \le X \le 1 + i\sqrt{1}) = P(X \le 1 + i) = F(1 + i) = 1 - e^{-(1)(1+i)}$.
   (c) The mean is $2 = 1/(1/2)$ and the variance is $4 = 1/(1/2)^2$ so we want the probabilities $P(0 \le X \le 2+i\sqrt{4}) = P(X \le 2+2i) = F(2+2i) = 1-e^{-(1/2)(2+2i)}$.

7. (a) The limits are correct at 0 and 1, so F is continuous. F is increasing on $[0, 1]$ and differentiable since $\sin^{-1}$ is increasing and differentiable. Thus F is a valid CDF. The pdf is $f(x) = \frac{1}{\pi}\frac{1}{\sqrt{x}\sqrt{1-x}}$
   (b) Even though f diverges at the endpoints 0 and 1, it is still integrable as an improper integral, so the probabilities are still finite.

8. (a) $F(X) = P(X \le x) = \int_0^x 12x^2(1 - x)dx = (4x^3 - 3x^4)|_0^x = 4x^3 - 3x^4$.
   (b) $P(0 < X < 1/2) = 4/8 - 3/16 = 5/16$.
   (c) $E(X) = \int_0^1 12x^3(1 - x)dx = 3x^4 - 12/5x^5 = 3/5$. $E(X^2) = \int_0^1 12x^4(1 - x)dx = 12/5x^5 - 2x^6 = 2/5$. $Var(X) = E(X^2) - E(X)^2 = 10/25 - 9/25 = 1/25$.

9. $\int_{-\infty}^{x} \frac{1}{\pi(1+x^2)}dx = \frac{1}{\pi}\tan^{-1}(x)|_{-\infty}^{x} = \tan^{-1}(x)/\pi + 1/2$.

2 Uniform and universality of the Uniform

10. (a) The union has total length $2+4=6$, so the desired probability is $\frac{6}{8}$.
    (b) The conditional distribution is $Unif(3,7)$.

11. (a) By location-scale, $E(U)=0$, $Var(U)=\frac{1}{3}$. Finally, $E(U^4)=\int_{-1}^{1} x^4/2dx = \frac{1}{10}+\frac{1}{10}=\frac{1}{5}$.
    (b) $P(U^2 \le x) = P(-\sqrt{x} \le U \le \sqrt{x}) = \sqrt{x}$ so this is not uniform. Then $f(x)=\frac{1}{2\sqrt{x}}$.

12. Let l denote the length of the stick, so the breakpoint is $U \sim Unif(0,l)$. The resulting lengths are U and $l-U$ and the length of the longer piece is $L=\max(U,l-U) \sim Unif(l/2,l)$. Thus the PDF is $2/l$ for $x \in [l/2,l]$ so the cdf is $2x/l-1$ and the average length is $3l/4$.

13. (a) We can reduce to the case when the breakpoint is $U \sim Unif(0,1/2)$, in which case $R=U/(1-U)$. Then $P(R<x)=P(U<(1-U)x)=P(U(1+x)<x)=P(U<x/(1+x))=2x/(1+x)$ for $0 \le x \le 1$. The pdf is $d/dx(x(1+x)^{-1}) = (1+x)^{-1}-x(1+x)^{-2}=2/(1+x)^2$.
    (b) $E(R)=\int_0^1 2x/(1+x)^2dx = 2log(2)-1 \approx .38629$
    (c) $E(1/R)=\int_0^1 2/(x(1+x)^2)dx=\infty$ so this expected value does not exist since it is infinite.

14. $P(X \le x) = P(U_1,\dots,U_n \le x) = \prod_i P(U_i \le x) = x^n \implies f(x)=nx^{n-1}$. $E(X)=\int_0^1 nx^n dx = \frac{n}{n+1}$.

15. $X=F^{-1}(U) \sim Expo(\lambda)$, where $F(x)=1-e^{-\lambda x}$ so $F^{-1}(u)=-\frac{1}{\lambda}\log(1-u)$.

16. (a) By LOTUS, $E(X^2)=\int_0^1 \log(u/(1-u))^2du$.
    (b) $E(X)=\int_0^1 \log(u)-\log(1-u)du = \int_0^1 \log(u)du - \int_0^1 \log(1-u)du = 0$ since u and $1-u$ are identically distributed.

17. $F^{-1}(u)=\sqrt[3]{-\log(1-u)}$ so $X=F^{-1}(U)$ has the desired distribution.

18. (a) $F(x)=P(X\le x)=\int_1^x a/x^{a+1}dx = -x^{-a}|_1^x = 1-x^{-a}$. This has all the properties of a CDF.
    (b) $F^{-1}(u)=(1-u)^{-1/a}$, so $F^{-1}(U)$ samples the Pareto distribution.

3 Normal

19. $Y=2Z+1$. $E(Y)=2E(Z)+1=1$ and $Var(Y)=4Var(Z)=4$.

20. (a) $\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt = \int_{-\infty}^{0}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt + \int_0^z \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt = 1/2+\frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt$. Make the substitution $u=t/\sqrt{2}$ with $du=dt/\sqrt{2}$ to continue the equality with $=1/2+\frac{1}{2}\frac{2}{\sqrt{\pi}}\int_0^{z/\sqrt{2}} e^{-u^2}du = 1/2+1/2erf(z/\sqrt{2})$.

(b) $erf(-z) = \frac{2}{\sqrt{\pi}} \int_0^{-z} e^{-x^2} dx = -\frac{2}{\sqrt{\pi}} \int_0^{z} e^{-y^2} dy = -erf(z)$.

21. (a) Constant multipliers do not affect the location of the inflection points, so we want to find where the second derivative of $e^{-z^2/2}$ is equal to 0. The first derivative is $-ze^{-z/2}$ and the second derivative is $-e^{-z^2/2} + z^2 e^{-z^2/2}$. Setting this equal to zero we find $z = \pm 1$ are the inflection points.

    (b) To get $N(\mu, \sigma^2)$, we use $\sigma Z + \mu$. The inflection points are at $\mu \pm \sigma$.

22. $P(-.4 < .2Z < .4) = P(-2 < Z < 2) = \Phi(2) - \Phi(-2)$. This is within two standard deviations of the mean, so the probability is $\approx .95$.

23. (a) If Alice sends a 0, then Bob understands assuming the noise is less than .5. If Alice sends a 1, then Bob understands assuming the noise is greater than $-.5$. Whether Alice sends a zero or one, the probability of understanding correctly is the same, so we only need to compute one of them. $P(\sigma Z < .5) = \Phi(.5/\sigma)$.

    (b) If σ is very small, then the probability goes to 1 which makes sense since if there is essentially no noise, then Bob can always understand correctly. If σ is very large, then the probability goes to $1/2$ which makes sense since if the noise overwhelms the actual signal, then Bob is basically guessing a random bit.

24. $P(0 < 8Z < 1) = P(0 < Z < 1/8) = \Phi(1/8) - 1/2$.

25. $P(X < Y) = P(X - Y < 0) = P(N(a - c, b + d) < 0) = P(\sqrt{b + d}Z + a - c < 0) = P(Z < (c - a)/\sqrt{b + d}) = \Phi((c - a)/\sqrt{b + d})$. If X and Y are iid, then we find $\Phi(0) = 1/2$. (In fact, as long as X and Y have the same mean, the result is 1/2.)

26. (a) $P(C < W) = P(C - W < 0) = P(N(c - w, \sigma^2 + 4\sigma^2) < 0) = P(\sqrt{5\sigma^2}Z + (c - w) < 0) = \Phi((w - c)/\sqrt{5\sigma^2})$.

    (b) $w - c > 0$.

    (c) $P(C < w + 10) > P(W < w + 10) \iff P(C - w - 10 < 0) > P(W - w - 10 < 0) \iff P(N(c - w - 10, 4\sigma^2) < 0) > P(N(-10, \sigma^2) < 0) \iff P(2\sigma Z + (c - w - 10) < 0) > P(\sigma Z - 10 < 0) \iff \Phi((w + 10 - c)/(2\sigma)) > \Phi(10/\sigma) \iff (w + 10 - c)/(2\sigma) > 10/\sigma \iff w - c > 10$.

27. No, by the symmetry property and the fact that the PDF is decreasing for $x > 0$ and increasing for $x < 0$, there is no other interval we can take. Also, the most probability is concentrated in that region so there is no chance.

28. $a(Y) = Y - 1.96\sigma, b(Y) = Y + 1.96\sigma$. Then $P(a(Y) < \mu < b(Y)) = P(Y - 1.96\sigma < \mu < Y + 1.96\sigma) = P(-1.96\sigma < \mu - Y < 1.96\sigma) = P(|\mu - Y| < 1.96\sigma) \approx .95$.

29. (a) $P(|X| < x) = 0$ if $x < 0$ and otherwise $P(|X| < x) = P(-x < X < x) = P(-x < \sigma Z + \mu < x) = P((-x - \mu)/\sigma < Z < (x - \mu)/\sigma) = \Phi((x - \mu)/\sigma) - \Phi((-x - \mu)/\sigma)$.

    (b) The PDF is $1/\sigma f((x - \mu)/\sigma) + 1/\sigma f((-x - \mu)/\sigma)$.

(c) The PDF is not continuous at zero, since the limit from the right is positive while the limit from the left is zero. This is not a problem since the function is integrable even if it is not continuous.

30. $P(SZ < z) = 1/2P(SZ < z|S = 1) + 1/2P(SZ < z|S = -1) = 1/2P(Z < z) + 1/2P(Z > -z) = 1/2\Phi(z) + 1/2\Phi(z) = \Phi(z)$ by symmetry of the uniform.

31. By universality of the uniform, $\Phi(Z) \sim Unif(0, 1)$, so $E(\Phi(Z)) = .5$.

32. (a) $P(1 \leq Z^2 \leq 4) = P(-2 < Z < -1) + P(1 < Z < 2) = (\Phi(2) - \Phi(-2)) - (\Phi(1) - \Phi(-1)) \approx .95 - .68 = .27$.

    (b) If $Z > t$, then the LHS is 1 and the RHS is strictly larger than 1. If $Z \leq t$, then the LHS is 0 and the RHS is 0.

    $\Phi(t) = P(Z \leq t) = 1 - P(Z > t) = 1 - E(I(Z > t)) \geq 1 - E(Z/tI(Z > t)) = 1 - \int_t^\infty z/t\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = 1 - 1/t\int_{t^2/2}^\infty \frac{1}{\sqrt{2\pi}}e^{-u}du = 1 - 1/(t\sqrt{2\pi})(-e^{-u})|_{t^2/2}^\infty = 1 - 1/(t\sqrt{2\pi})(e^{t^2/2}) = 1 - \varphi(t)/t$.

33. (a) $E(Z^4) = \int_{-\infty}^\infty z^4\varphi(z)dz = \int_0^\infty z^2g(z)dz$.

    (b) $E(Z^2 + Z + \Phi(Z)) = E(Z^2) + E(Z) + E(\Phi(Z)) = Var(Z) + E(Z)^2 + E(Z) + E(Unif(0, 1)) = 1 + 0 + 0 + .5 = 1.5$.

34. $E(X) = E(ZI_{Z>0}) = \int_0^\infty z\varphi(z)dz\frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-u}du = \frac{1}{\sqrt{2\pi}} - e^{-u}|_0^\infty = \frac{1}{\sqrt{2\pi}}$.

    $Var(X) = E(Z^2I_{z>0}) - 1/(2\pi) = 1/2(1 - 1/\pi)$.

35. $E((Z-c)I(Z-c > 0)) = \int_c^\infty(z-c)\varphi(z)dz = \int_c^\infty z\varphi(z)dz - c\int_c^\infty \varphi(z)dz = \frac{1}{\sqrt{2\pi}}e^{-c^2/2} - c(1 - \Phi(c)) = \varphi(c) - c(1 - \Phi(c))$.

4 Exponential

36. (a) Alice waits $Expo(2\lambda)$ until she is served by a Clerk. However, by the memoryless property of the exponential distribution, the time until Alice is served as well as the time until the remaining of Bob and Claire is served are independent $Expo(\lambda)$ random variables, so the probability that Alice is last is the same as the probability that Alice is second to last, and both are 1/2.

    (b) $E(X_2 \sim Expo(2\lambda) + X \sim Expo(\lambda)) = \frac{1}{2\lambda} + \frac{1}{\lambda}$.

37. (a) Find t such that $F(t) = P(T < t) = .5$. $F(t) = 1 - e^{-\lambda t}$ with inverse $F^{-1}(u) = -\frac{1}{\lambda}\log(1 - u)$.

    (b) $P(t < T < t + \varepsilon|t < T) = P(T < \varepsilon) = 1 - e^{-\lambda\varepsilon} = 1 - (1 - \lambda\varepsilon) = \lambda\varepsilon$.

    (c) $L = \min(T_1, \ldots, T_n) \sim Expo(n\lambda)$, so $P(L \leq t) = 1 - e^{-n\lambda t}$. $E(L) = \frac{1}{n\lambda}$ and $Var(L) = \frac{1}{n^2\lambda^2}$.

(d) By memorylessness, $M \sim Expo(n\lambda) + Expo((n-1)\lambda) + \cdots + Expo(\lambda)$. Thus $E(M) = \frac{1}{n\lambda} + \frac{1}{(n-1)\lambda} + \cdots + \frac{1}{\lambda}$. These are disjoint increments, so they are independent, and thus $Var(M) = \frac{1}{n^2\lambda^2} + \frac{1}{(n-1)^2\lambda^2} + \cdots + \frac{1}{\lambda^2}$.

38. (a) $P(O > 18k) = 1 - F(18k) = 1 - (1 - e^{-18k/12k}) = e^{-3/2}$ so the number of offers Fred will have is distributed as first success with parameter $p = e^{-3/2}$, so the expected number of offers is $e^{3/2}$.

    (b) $E(O|O > 18k) = \int_{18k}^{\infty} x(1/12k)e^{-x/12k}e^{3/2}dx = 30000$. For a cleverer argument, the memoryless property implies that $O|O > 18k \sim 18k + O$, so $E(O|O > 18k) = 18k + E(O) = 30k$.

39. This is the max of three Expo(1/12k) distributions, which has expected value $4k + 6k + 12k = 22k$.

40. (a) By memorylessness, the time to the next bus is distributed as $Expo(10)$. Also by memorylessness, the time to the previous bus is also $Expo(10)$, so overall the expected length of time Fred sees between buses is 20 minutes.

    (b) Consider a distribution where the buses either arrive in 1 second or 2 hours. It is very unlikely that Fred arrives in the one second interval between bus arrivals, so he will almost always see the 2 hour interval between buses.

41. (a) By memorylessness, Fred will have to wait $1/\lambda_1$ minutes for his bus.

    (b) The hint is somewhat intuitive: the time until any bus arrival has mean $1/(\lambda_1+\lambda_2)$ since it is the minimum of $Expo(\lambda_1)$ and $Expo(\lambda_2)$. The ratio of those arrivals which are route 1 is λ_1, so the probability that the first arrival is route 1 is $\lambda_1/(\lambda_1 + \lambda_2)$.

    By memorylessness, it suffices to find the probability that a route 1 bus will arrive before a route 2 bus, since if the route 1 bus arrives, then we are again in the situation of waiting for either a route 1 bus or a route 2 bus. Thus the probability that n route 1 buses arrive before the first route 2 bus is $(\lambda_1/(\lambda_1 + \lambda_2))^n$.

    (c) This is the max of two independent exponentials, so $1/2\lambda + 1/\lambda$.

42. $P(X \leq k) = P(T \leq (k+1)/365) = 1 - e^{-1/5*(k+1)/365}$, so $P(X = k) = P(X \leq k) - P(X \leq k-1) = (1 - e^{-1/5*(k+1)/365}) - (1 - e^{-1/5*k/365}) = e^{-1/5*k/365} - e^{-1/5*(k+1)/365} = e^{-1/5*k/365}(1 - e^{-1/5*1/365}) = e^{-k/N}(1 - e^{-1/N})$, where $N = 5 * 365$. This is a geometric distribution with sucess parameter $1 - e^{-1/N}$.

43. (a) $G\Delta t = T$

    (b) $P(T > t) = P(G > t/\Delta t)$, so there must at least $t/\Delta t$ failures, which has probability $(1 - \lambda\Delta t)^{t/\Delta t}$ and thus $P(T \leq t) = 1 - (1 - \lambda\Delta t)^{t/\Delta t}$. The limit as $\Delta t \to 0$ is $1 - e^{-\lambda t}$ which is exactly the $Expo(\lambda)$ CDF.

44. (a) For positive x, the Laplace distribution PDF is half the exponential PDF with parameter $\lambda = 1$. Then rather than ignoring the negative numbers, the Laplace distribution "symmetrizes" the PDF by reflecting it over the Y-axis.

(b) $P(SX < x) = 1/2P(X < x) + 1/2P(X > -x)$. If x is negative, then $P(X < x) = 0$ and $P(X > -x) = P(X > |x|) = 1 - P(X < |x|) = 1 - F(|x|) = e^{-|x|}$ so in this case $P(SX < x) = 1/2e^{-|x|} = 1/2e^{x}$. If x is positive, then $P(X < x) = F(x)$ and $P(X > -x) = 1$, so in this case $P(SX < x) = 1/2(1 - e^{-x}) + 1/2 = 1 - 1/2e^{-x}$. Thus, the derivative is $f(x) = 1/2e^{x}$ if x is negative and $1/2e^{-x}$ if x is positive, which coincides exactly with the Laplace PDF.

45. $P(T > .1) = P(N_{.1} < 3) = P(Pois(.1 * 20) < 3) = e^{-2}(1 + 2 + 4/2) = 5e^{-2}$.

46. (a) The probability of surviving until time t is $1 - F(t) = P(T \geq t)$. Then the probability density of death at time t is the probability that $t \leq T \leq t + \Delta t$ as $\Delta t \to 0$, which is the derivative of the CDF at time t so the hazard function is the probability density of death at time t.

    (b) $h(t) = \lambda e^{-\lambda t}/(1 - (1 - e^{-\lambda t})) = \lambda$. Conversely, $\lambda = F'(t)/(1 - F(t)) \implies -\log(1 - F(t)) = \lambda t \implies F(t) = 1 - e^{-\lambda t}$ is the exponential CDF.

47. (a) $F'/(1 - F) = h \implies -\log(1 - F) = \int_0^t h(t)dt \implies F = 1 - \exp(-\int_0^t h(t)dt)$.

    (b) Take a derivative of both sides of the previous equation to get the desired result.

48. $E(X^3) = \int_0^\infty x^3(\lambda e^{-\lambda x})dx = x^3(-e^{-\lambda x})|_0^\infty - \int_0^\infty 3x^2(-e^{-\lambda x})dx = \frac{3}{\lambda}\int_0^\infty x^2\lambda e^{-\lambda x}dx$.
On the other hand, $E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}dx = Var(X) + E(X)^2 = 2/\lambda^2$ so the final result is $6/\lambda^3$. Looks like the general result will be $n!/\lambda^n$.

49. (a) The support for the Gumbel distribution is the whole real line. $P(-\log X < x) = P(X > e^{-x}) = e^{-e^{-x}}$ which has the correct limiting behavior as $x \to -\infty$ and $x \to \infty$.

    (b) $M_n \sim Expo(n\lambda) + Expo((n-1)\lambda) + \cdots + Expo(\lambda)$.
    $P(M_n - \log n < x) = \prod_i P(X_i < x + \log n) = (1 - e^{-(x+\log n)})^n = (1 - e^{-x}/n)^n \approx e^{-e^{-x}}$ as $n \to \infty$.

5 Mixed practice

50. If X and Y are iid, then there's nothing to distinguish between them, so the probability that one is larger is the same as the probability that the other is larger. If they are not iid, then it may be the case that $X > Y$ with probability 1.

51. (a) $Var(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2 \leq E(X) - \mu^2 = \mu - \mu^2$ since $X^2 \leq X$. Since $0 \leq X \leq 1$, then $0 \leq \mu \leq 1$ so $0 \leq \mu(1 - \mu) \leq 1/4$ since $\mu - \mu^2$ is a parabola with its maximum at $\mu = 1/2$.

    (b) In order for $Var(X) = 1/4$, then we must have $E(X^2) = \mu = E(X)$. Then $E(X^2 - X) = E(X(1 - X)) = 0$, so $X(1 - X) = 0$ with probability 1, meaning that $X = 1$ or $X = 0$ with probability 1. Thus X must be a Bernoulli trial with probability $p = 1/2$ of success.

52. (a) $E(X) = \int_0^\infty x^2 e^{-x^2/2}dx$. $\varphi = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ and $E(Z^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty z^2 e^{-z^2/2}dz = 1 \implies \int_{-\infty}^\infty z^2 e^{-z^2/2}dz = \sqrt{2\pi} \implies \int_0^\infty z^2 e^{-z^2/2}dz = \sqrt{\pi/2}$ by symmetry of the integrand.

(b) $E(X^2) = \int_0^\infty x^2(xe^{-x^2/2})dx = x^2(-e^{-x^2/2})|_0^\infty - \int_0^\infty 2x(-e^{-x^2/2})dx = 2\int_0^\infty xe^{-x^2/2}dx = 2$ since the Rayleigh pdf must integrate to 1.

Alternatively, $E(X^2) = \int_0^\infty x^2(xe^{-x^2/2})dx = \int_0^\infty 2ue^{-u}du = 2E(Expo(1)) = 2$.

53. $E((T-\theta)^2) = Var(T-\theta) + (E(T-\theta))^2 = Var(T) + b(T)^2$.

54. (a) Naively, one might guess 11 percent (1/9) if the first digit is uniform on 1-9. However, it feels like 1 might be a little more common than expected, so I'll guess 20 percent.

(b) The data lines up fairly closely with Benson's law.

(c) $\int_1^{10} c/xdx = c\ln(x)|_1^{10} = c\ln(10) = 1 \implies c = 1/\ln(10)$.

$P(Y < y) = P(X < y/a) = \int_1^{y/a} c/xdx$ with derivative c/y. The domain changes to $a < Y < 10a$ with the same constant c as before.

More interesting interpretation: WLOG we may assume that $1 < a \leq 10$ by rescaling by powers of 10. Then if $1 < X < \frac{10}{a}$, we have $1 < Y < 10$. Else if $\frac{10}{a} < X < \frac{100}{a}$, then $1 < \frac{Y}{10} < 100$.

Then for $1 < x \leq 10$, $P(Y < x) = P(aX < x|1 < X < \frac{10}{a})P(1 < X < \frac{10}{a}) + P(aX < x|\frac{10}{a} < X < \frac{100}{a})P(\frac{10}{a} < X < \frac{100}{a}) = P(X < (x/a)|1 < X < \frac{10}{a})P(1 < X < \frac{10}{a}) + P(aX < x|\frac{10}{a} < X < \frac{100}{a})P(\frac{10}{a} < X < \frac{100}{a})$.

(d) $P(j < X < j+1) = \int_j^{j+1} c/xdx = c\ln(x) = c\ln((j+1)/j) = \log_{10}((j+1)/j)$.

55. (a) Let $p = P(N(0,4) > 4) = P(2Z > 4) = P(Z > 2) = 1 - \Phi(2)$. Then j has the first success distribution with parameter p so $E(J) = \frac{1}{1-\Phi(2)}$.

(b) $E(R) = \int_{-\infty}^\infty g(x)/f(x)f(x)dx = \int_{-\infty}^\infty g(x)dx = 1$ since $g(x)$ is a PDF.

(c) This is the CDF of the Gumbel distribution. By universality of the uniform, $W \sim Unif(0,1)$ so $E(W) = 1/2$ and $Var(W) = 1/12$.

56. (a) $E(Z^2\Phi(Z)) = \int_{-\infty}^\infty z^2\Phi(z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$

(b) $P(\Phi(Z) < 2/3) = P(U < 2/3) = 2/3$ by universality of the uniform.

(c) $P(X < Y < Z) = \frac{1}{6}$ since there are 6 possible permutations of X, Y, Z and any of them is equally likely since X, Y, Z are iid.

57. (a) $U = \Phi(Z) \sim Unif(0,1)$, so $W = U^2$ and for $0 \leq x \leq 1, P(W \leq x) = P(U \leq \sqrt{x}) = \sqrt{x}$. Then $f_W(x) = \frac{1}{2\sqrt{x}}$.

(b) $E(W^3) = \int_0^1 x^3 f_W(x)dx = \int_{-\infty}^\infty \Phi(x)^6\varphi(x)dx$.

(c) Using that X, Y, Z are iid, so their sum is again normal, $P(X + 2Y - 2Z < 3) = P(N(0,9) < 3) = P(3Z < 3) = P(Z < 1) = \Phi(1)$.

58. (a) $E(Y) = \int_{-\infty}^\infty \max(z,0)\varphi(z)dz = \int_0^\infty z\varphi(z)dz = 1/\sqrt{2\pi}(-e^{-u}|_0^\infty) = 1/\sqrt{2\pi}$.

(b) $P(Y = 0) = 1/2$ so N is first success with parameter $1/2$. Thus $E(N) = 2$.

(c) $P(Y < y) = 0$ if $y < 0$ and $P(Y < y) = \Phi(y)$ if $y > 0$.

59. (a) The average length of a random arc (viewing the arc as randomly choosing one of the three) is $2\pi/3$, since we can set up indicator variables for the length of each arc I_A, I_B, I_C. Since they are uniformly chosen, then they are identically distributed and $E(I_A + I_B + I_C) = E(2\pi) = 2\pi \implies E(I_A) = E(I_B) = E(I_C) = 2\pi/3$. However, L does not refer to a random arc, so $L \neq I_A, I_B, I_C$, and this argument would only work if all arc lengths are equally likely.

(b) Cutting the circle at $(1, 0)$, we can view the three points as three iid $Unif(0, 2\pi)$ random variables. Then the closest points to $(1, 0)$ are the min and max, and the corresponding lengths are $\min(A, B, C)$ and $2\pi - \max(A, B, C) = \min(2\pi - A, 2\pi - B, 2\pi - C)$ which is identically distributed with $\min(A, B, C)$ since $2\pi - Unif(0, 2\pi) \sim Unif(0, 2\pi)$.

Then $P(\min(A, B, C) > x) = P(A, B, C > x) = P(Unif(0, 2\pi) > x)^3 = 1 - (1 - x/2\pi)^3$ and the PDF is $3/2\pi(1 - x/2\pi)^2$. The min of n standard uniform random variables has mean $1/(n+1)$, so by scale transform the min of these three $Unif(0, 2\pi)$ has mean $2\pi/4 = \pi/2$.

(c) $E(L) = \pi$.

60. (a) Let I_j be the indicator of whether jumper j is the best in recent memory. Then $P(I_j = 1) = 1/3$ so $E(\sum_j I_j) = (n - 3 + 1)/3$.

(b) $P(A_3) = P(A_4) = 1/3$. $P(A_3 \cap A_4) = 2/4! = 1/12 \neq 1/9$. Thus the two events are not independent.

61. (a) Since T arriving before C is equally likely as C arriving before T, it doesn't matter which actually occurs. Thus, we are interested in the number of people arriving between C and T. Let I_j be the indicator of whether person j arrives when the party is fun. Since person j is equally likely to arrive before either, between both, or after both, then $P(I_j = 1) = 2/3! = 1/3$ so $E(\sum_j I_j) = n/3$.

(b) $P(J) = P(R) = 1/3$. $P(JR) = 4/4! = 1/3! \neq 1/9$, so these events are not independent.

(c) The two events are not independent since if J arrives during a fun time, it makes it more likely that the fun time is long so R is also likely to arrive during a fun time.

They are not conditionally independent for essentially the same reason. Suppose there are f fun "slots" where J or R could arrive. There are $n+1$ "slots" total, f of which are "fun" (same for both), so both have probability $f/(n+1)$ of arriving during a fun time.

The number of ways to place them both is $2(n+1) + (n+1) * n = (n+1)(n+2)$ (either they are adjacent arrivals and we need to pick who is first or they are not adjacent arrivals and we need to pick who is first). The number of ways to place them in fun "slots" is $2f + f(f - 1) = f(f + 1)$. Thus the probability is

$f(f+1)/(n+1)(n+2) \neq f^2/(n+1)^2$ (although they are close if f and n are sufficiently large.

62. (a) Let I_j be the indicator of year j being record high or record low. Then $P(I_j = 1) = P(H_j) + P(L_j) = 2/j$, so $E(\sum_j I_j) = \sum_j 1/j$.

    (b) Let I_j be the indicator of record low followed by record high. Then $P(I_j) = 1/j * 1/(j+1)$, so $E(\sum_j I_j) = \sum_j 1/(j(j+1))$.

    (c) $N > n$ means that of the first n years, year 1 was still the record high. This has probability $1/n$. The PMF of N is $P(N = n) = P(N > (n-1)) - P(N > n) = 1/(n-1) - 1/n = 1/n(n-1)$.

    (d) $E(N) = \sum_j j/j(j-1) = 1/(j-1)$ which diverges.