Chris Piech
CS109

Section #3
October 13, 2017

Section #3 Solutions

1. **Website Visits**: Let X be the number of minutes that a user stays. $X \sim \text{Exp}(\lambda = \frac{1}{5})$.

$$\begin{aligned} P(X > 10) &= 1 - F_X(10) \\ &= 1 - (1 - e^{\lambda 10}) = e^{-2} \approx 0.1353 \end{aligned}$$

2. **Approximating Normal**: The number of users that log in B is binomial: $B \sim \text{Bin}(n = 100, p = 0.2)$. It can be approximated with a normal that matches the mean and variance. Let C be the normal that approximates B.

$E[B] = np = 20.$

$Var(B) = np(1-p) = 16$

$C \sim N(\mu = 20, \sigma^2 = 16).$

$$\begin{aligned} P(B > 21) &\approx P(C > 21.5) \\ &= P\Big(\frac{C-20}{\sqrt{16}} > \frac{21.5-20}{\sqrt{16}}\Big) \\ &= P(Z > 0.375) \\ &= 1 - P(Z < 0.375) \\ &= 1 - \phi(0.375) = 1 - 0.6462 = 0.3538 \end{aligned}$$

3. **Continuous Random Variable**:

a. We need $\int_{-\infty}^{\infty} dx\, f_X(x) = 1$.

$$\begin{aligned} 1 = \int_{-\infty}^{\infty} f_X(x)dx &= \int_0^1 c(e^{x-1} + e^{-x})dx \\ &= c\left[e^{x-1} - e^{-x}\right]_{x=0}^1 \\ 1 &= c(e^{1-1} - e^{-1} - (e^{0-1} - e^{-0})) \\ c &= \frac{1}{1 - e^{-1} - (e^{-1} - 1)} = \frac{1}{2 - \frac{2}{e}} \end{aligned}$$

b.

$$\begin{aligned} P(X > 0.75) &= \int_{0.75}^1 dx\, c(e^{x-1} + e^{-x}) \\ &= -c\left[e^{x-1} - e^{-x}\right]_{x=0.75}^1 \\ &= \boxed{-c\left(e^{1-1} - e^{-1} - (e^{0.75-1} - e^{-0.75})\right)} \\ &= -c\left(1 - e^{-1} - e^{-0.25} + e^{-0.75}\right) = \frac{1 - e^{-1} - e^{-0.25} + e^{-0.75}}{2 - \frac{2}{e}} \end{aligned}$$

4. Student Heights

a. $P(\text{height} > 72) = 1 - P(\text{height} < 72) = 1 - \Phi(\frac{72-65}{3.5}) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$

b. $\Phi(\frac{x-\mu}{\sigma}) = 0.1$. From the reverse phi table lookup, we get $\frac{x-\mu}{\sigma} = -1.29$. Thus $-1.29 = \frac{x-\mu}{\sigma} = \frac{65-70}{\sigma}$ and $\sigma = 3.88$.

c. P(at least one man taller than Chris) = 1 − P(all men shorter). P(any one man in the section is shorter = $\Phi(\frac{77-70}{3.88}) = \Phi(1.80) = 0.9641$. $1 - P(\text{all men shorter}) = 1 - (0.9641)^6 = 0.803$.
P(height = 77") = 0 because we cannot talk about precise equalities when dealing with continuous distributions.
P(76.5" < height < 77.5") = $\Phi(\frac{77.5-70}{3.88}) - \Phi(\frac{76.5-70}{3.88}) = \Phi(1.933) - \Phi(1.675) = 0.973 - 0.953 = 0.020$

5. Elections

a. $P(\text{random person votes for A}) = \frac{votes for A}{total votes} = \frac{4881}{7453} = 0.655$
Now, let X be the number of votes for candidate A. We assume that X $\sim Bin(42495933.7, 0.655)$.

- Since n is so large, we can approximate X using a normal Y $\sim N(np, np(1-p))$.
- $\mu = np = 42495933.7$, Variance = $np(1-p) = 14665173.9$ Std Dev = 3829.5
- Votes to win = $\frac{64888792}{2} = 32444396$
- P(A gets enough votes) = $P(X > 32444396) = P(Y > 32444395.5)$ (using a continuity correction) = 1.00

b. $S_A \sim N(5.324, 16.436)$
$S_B \sim N(2.926, 16.436)$

$P(S_A > S_B) \approx 0.60$

Algorithm (a) makes very few assumptions, and simplicity can be useful, but it does assume that each voter is independent - which we definitely know isn't the case in real elections. Algorithm (b) allows us to model bias (using the weights we incorporated), and doesn't think of each voter as necessarily independent.