Probability Distributions for Discrete Random Variables

Ben Goodrich

January 31, 2019

Obligatory Disclosure

- Ben is an employee of Columbia University, which has received several research grants to develop Stan
- Ben is also a manager of GG Statistics LLC, which uses Stan for business purposes
- According to Columbia University policy, any such employee who has any equity stake in, a title (such as officer or director) with, or is expected to earn at least $\$5,000.00$ per year from a private company is required to disclose these facts in presentations

Bernoulli Distribution

- Tuesday we talked about the Bernoulli distribution over $\Omega = \{0, 1\}$, whose Probability Mass Function (PMF) is $\Pr\left(x | \pi\right) = \pi^x (1 - \pi)^{1-x}$, which depends on a possibly unknown probability parameter $\pi \in [0, 1]$
- $\mathbb{E}X = 0 \times (1 - \pi) + 1 \times \pi = \pi$
- $\mathbb{E}(X - \pi)^2 = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 = \mathbb{E}X - \pi^2 = \pi - \pi^2 = \pi(1 - \pi)$
- You could write a model where π is a function of predictors for each observation, as in $\pi(z) = \frac{1}{1+e^{-\alpha-\beta z}}$ for a logit model

Binomial Distribution

- A Binomial random variable can be defined as the sum of n INDEPENDENT Bernoulli random variables with the same π. What is Ω in this case?
- What is an expression for the expectation of a Binomial random variable?
- What is an expression for the variance of a Binomial random variable?
- What is an expression for the PMF, $\Pr\left(x \middle| \pi, n = 3\right)$? Hint: 8 cases to consider
- All succeed, π^3 or all fail, $(1-\pi)^3$
- 1 succeeds and 2 fail $\pi^1(1-\pi)^{3-1}$ but there are 3 ways that could happen
- 2 succeed and 1 fails $\pi^2(1-\pi)^{3-2}$ but there are 3 ways that could happen
- In general, $\Pr\left(x \middle| n, \pi\right) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$, where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ is the number of ways x can occur in n tries (if order is unimportant) where ! indicates the factorial function

Back to Bowling

- Why is the binomial distribution with $n = 10$ inappropriate for the first roll of a frame of bowling?
- Could the Bernoulli distribution be used for success in getting a strike?
- Could the Bernoulli distribution be used for the probability of knocking over the frontmost pin?
- If $X_i = \mathbb{I}\{\text{pin i is knocked down}\}$ and π_i is the probability in the i-th Bernoulli distribution, what conceptually is

$$\Pr\left(x_1 \middle| \pi_1\right) \prod_{i=2}^{10} \Pr\left(x_i \middle| \pi_i, X_1 = x_1, X_2 = x_2, \ldots, X_{i-1} = x_{i-1}\right)?$$

Poisson Distribution for Counts

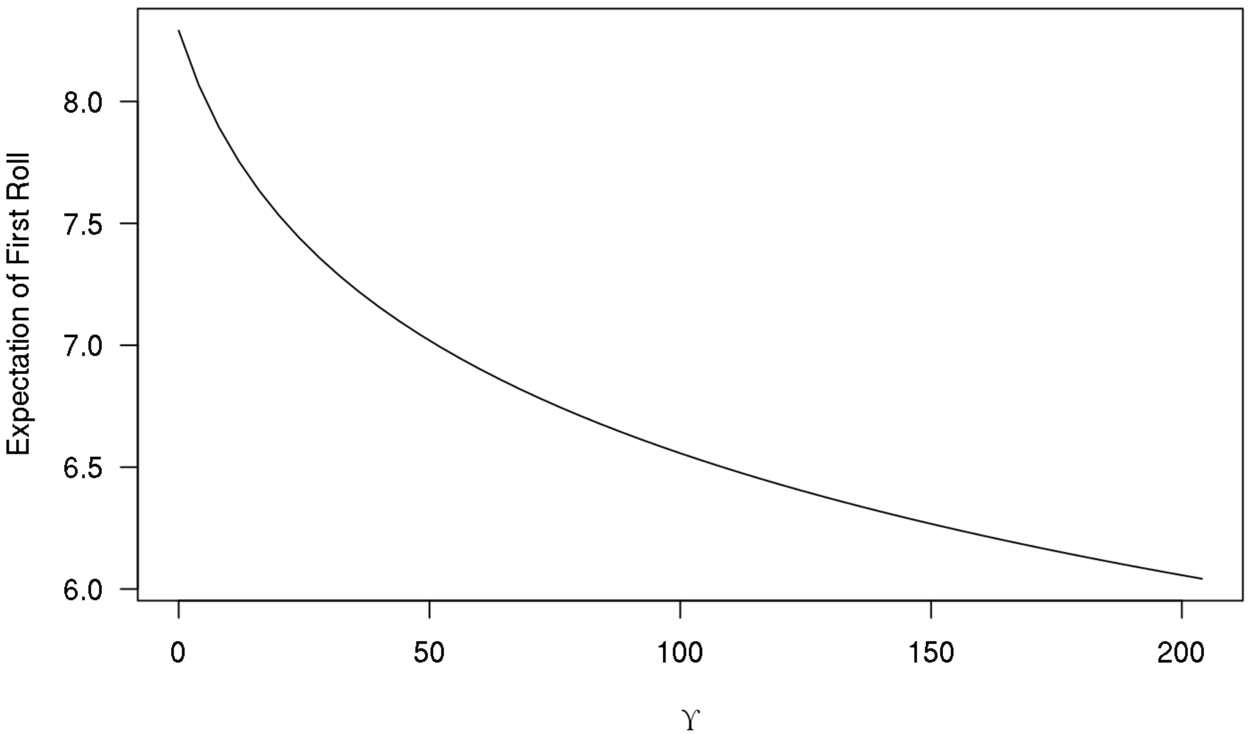
- Let $n \uparrow \infty$ and let $\pi \downarrow 0$ such that $\mu = n\pi$ remains fixed and finite. What happens to the binomial PMF, $\Pr(x | n, \pi) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$?
  - $(1-\pi)^{n-x} = \left(1 - \frac{\mu}{n}\right)^{n-x} = \left(1 - \frac{\mu}{n}\right)^{n} \times \left(1 - \frac{\mu}{n}\right)^{-x} \to e^{-\mu} \times 1$
  - $\binom{n}{x} \pi^x = \frac{n!}{x!(n-x)!} \frac{\mu^x}{n^x} = \frac{n \times (n-1) \times (n-2) \times \cdots \times (n-x+1)}{n^x} \frac{\mu^x}{x!} \to 1 \times \frac{\mu^x}{x!}$
  - Thus, $\Pr(x | \mu) = \frac{\mu^x e^{-\mu}}{x!}$ is the PMF of the Poisson distribution over $\Omega = \{0, \mathbb{Z}_+\}$, which is a common distribution for count variables
- What is the variance of a Poisson random variable?

Parameterized Bowling

- Bell numbers are defined as $\mathcal{B}_0 = 1$, $\mathcal{B}_1 = 1$, and else $\mathcal{B}_{n+1} = \sum_{k=0}^{n} \binom{n}{k} \mathcal{B}_k$
- Let $\Pr\left(x \middle| n, \Upsilon\right) = \frac{\binom{n+\Upsilon}{x+\Upsilon}\mathcal{B}_{x+\Upsilon}}{\mathcal{B}_{n+1+\Upsilon} - \sum_{k=0}^{\Upsilon-1}\binom{n+\Upsilon}{k}\mathcal{B}_k}$ where $\Upsilon \in \{0, \mathbb{N}_+\} = \Theta$ is an unknown
- Use a Poisson distribution with expectation μ to represent beliefs about Υ
- Can update those beliefs with data on pins knocked down in bowling

```
B <- Vectorize(memoise::memoise(numbers::bell)) # makes it go faster
Pr <- function(x, n = 10, Upsilon = 0) { # probability of knocking down x out of n pins
  stopifnot(length(n) == 1, is.numeric(x), all(x == as.integer(x)))
  numer <- B(x + Upsilon) * choose(n + Upsilon, x + Upsilon)
  denom <- B(n + 1 + Upsilon) # ^^^ choose(n, k) is n! / (k! * (n - k)!)
  if(Upsilon > 0) denom <- denom -
    sum(choose(n + Upsilon, 0:(Upsilon - 1)) * B(0:(Upsilon - 1)))
  return(ifelse(x > n, 0, numer / denom))
}
```

How to Select μ in the Poisson Prior?



Using Bayes Rule with Bowling Data

```
frames <- cbind(x_1 = c(9, 8, 10, 8, 7, 10, 9, 6, 9),
                x_2 = c(1, 1,  0, 2, 2,  0, 0, 3, 0))
```

- Suppose that you observe the first $J = 9$ frames of bowling on the same person. Your posterior beliefs about Υ are given by

$$\Pr\left(\Upsilon \mid x_{1,1}, x_{1,2}, \dots, x_{J,1}, x_{J,2}, \mu\right) = \frac{\Pr\left(\Upsilon \mid \mu\right) \Pr\left(x_{1,1}, x_{1,2}, \dots, x_{J,1}, x_{J,2} \mid \Upsilon\right)}{\Pr\left(x_{1,1}, x_{1,2}, \dots, x_{J,1}, x_{J,2}\right)}$$

- $\Pr\left(\Upsilon \mid \mu\right) = \frac{\mu^{\Upsilon} e^{-\mu}}{\Upsilon!}$ is the PMF for a Poisson distribution
- What is $\Pr\left(x_{1,1}, x_{1,2}, \dots, x_{J,1}, x_{J,2} \mid \Upsilon\right)$?
- Due to the independence of frames, $\Pr\left(x_{1,1}, x_{1,2}, \dots, x_{J,1}, x_{J,2} \mid \Upsilon\right) = \prod_{j=1}^{J} \Pr\left(x_{j,1} \mid n = 10, \Upsilon\right) \Pr\left(x_{j,2} \mid n = 10 - x_{j,1}, \Upsilon\right)$
- What is $\Pr\left(x_{1,1}, x_{1,2}, \dots, x_{J,1}, x_{J,2}\right)$?

Denominator of Bayes Rule

- $\Pr\left(x_{1,1}, x_{1,2}, \ldots, x_{J,1}, x_{J,2}\right) = \mathbb{E}g\left(\Upsilon\right)$ w.r.t. the prior PMF $\Pr\left(\Upsilon \middle| \mu\right)$, where $g\left(\Upsilon\right) = \Pr\left(x_{1,1}, x_{1,2}, \ldots, x_{9,1}, x_{9,2} \middle| \Upsilon\right)$. $g\left(\Upsilon\right)$ called the LIKELIHOOD function of Υ (evaluated at the observed data)

- The expected likelihood can be computed in this case as

$$\mathbb{E}g\left(\Upsilon\right) = \sum_{i=0}^{\infty} \Pr\left(i \middle| \mu\right) \Pr\left(x_{1,1}, x_{1,2}, \ldots, x_{J,1}, x_{J,2} \middle| \Upsilon = i\right) =$$

$$\sum_{i=0}^{\infty} \frac{\mu^i e^{-\mu}}{i!} \prod_{j=1}^{J} \Pr\left(x_{j,1} \middle| n = 10, \Upsilon = i\right) \Pr\left(x_{j,2} \middle| n = 10 - x_{j,1}, \Upsilon = i\right)$$

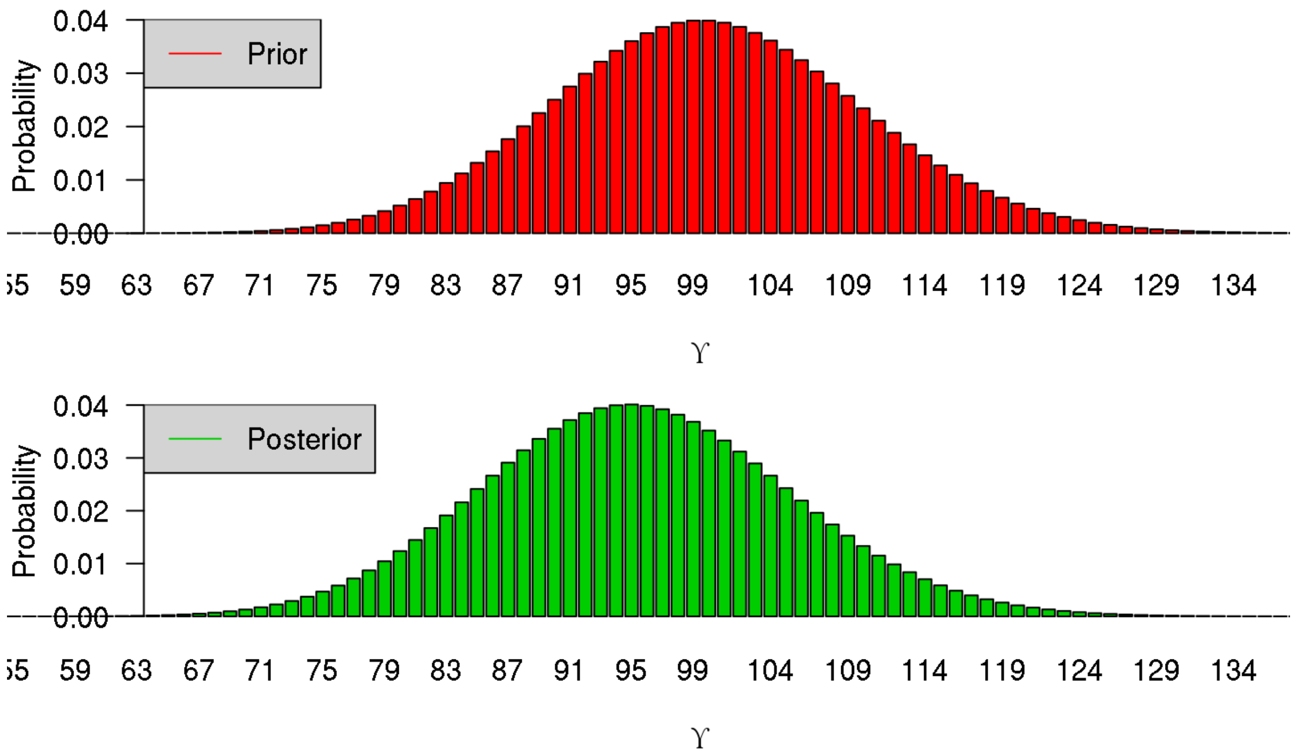
- In practice, when evaluating an infinite sum we just stop once the terms get close enough to zero

Calculating Posterior Probabilities in R

```
mu <- 100                 # for example, can be any positive real number
Theta <- 0:207            # 207 is the biggest a laptop can handle
prior <- dpois(Theta, mu) # dpois() is the Poisson PMF
numer <- prior * sapply(Theta, FUN = function(i) { # sapply applies the given function
  J <- nrow(frames)                                # to each element of the first argument
  Pr_game <- Pr(frames[ , "x_1"], n = 10, Upsilon = i) * sapply(1:J, FUN = function(j)
             Pr(frames[j, "x_2"], n = 10 - frames[j, "x_1"], Upsilon = i))
  prod(Pr_game)
})
post <- numer / sum(numer)

barplot(prior, names.arg = Theta, col = 2, ylim = range(post))
barplot(post, names.arg = Theta, col = 3)
```

Comparison of Prior and Posterior Probabilities



Hypergeometric Distribution

- The binomial distribution corresponds to sampling WITH replacement
- The hypergeometric distribution corresponds to sampling WITHOUT replacement and has PMF $\Pr\left(x \middle| N, K, n\right) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$ where
  - N is the (finite) size of the set being drawn from
  - K is the number of successes in that finite set
  - n is the number of times you draw without replacement
- What is the probability of drawing two cards from a deck with the same value?
- Intuitively, the probability of any pair should be $\frac{4}{52} \times \frac{3}{51} = \frac{1}{221}$ and there are 13 ways to do that so `13 * dhyper(x = 2, 4, 52 - 4, 2)`