

# Bayes Rule with Continuous Random Variables

Ben Goodrich

February 07, 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
- I do not have any connection to Mapp Biopharmaceutical, which we are going to use as an example today

# Beta Distribution

- The Beta distribution for  $\pi \in \Theta = [0, 1]$  has two positive shape parameters  $a$  and  $b$  and its PDF involves the Beta function  $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ :

$$f(\pi|a, b) = \frac{1}{B(a, b)} \pi^{a-1} (1 - \pi)^{b-1} \propto \pi^{a-1} (1 - \pi)^{b-1}$$

- Its expectation is  $\mu = \frac{a}{a+b}$  & mode is  $M = \frac{a-1}{a+b-2}$  but only exists if  $a, b > 1$
- Its median,  $m \approx \frac{a - \frac{1}{3}}{a+b - \frac{2}{3}}$ , always exists but approximation assumes  $a, b > 1$
- Given  $M, m \in (0, 1)$ , you can [solve](#) for  $a > 1$  and  $b > 1$ 
  - $a = \frac{m(4M-3)+M}{3(M-m)}$  while  $b = \frac{m(1-4M)+5M-2}{3(M-m)}$
  - But  $m$  must be between  $\frac{1}{2}$  and  $M$  in order for  $a > 1$  and  $b > 1$

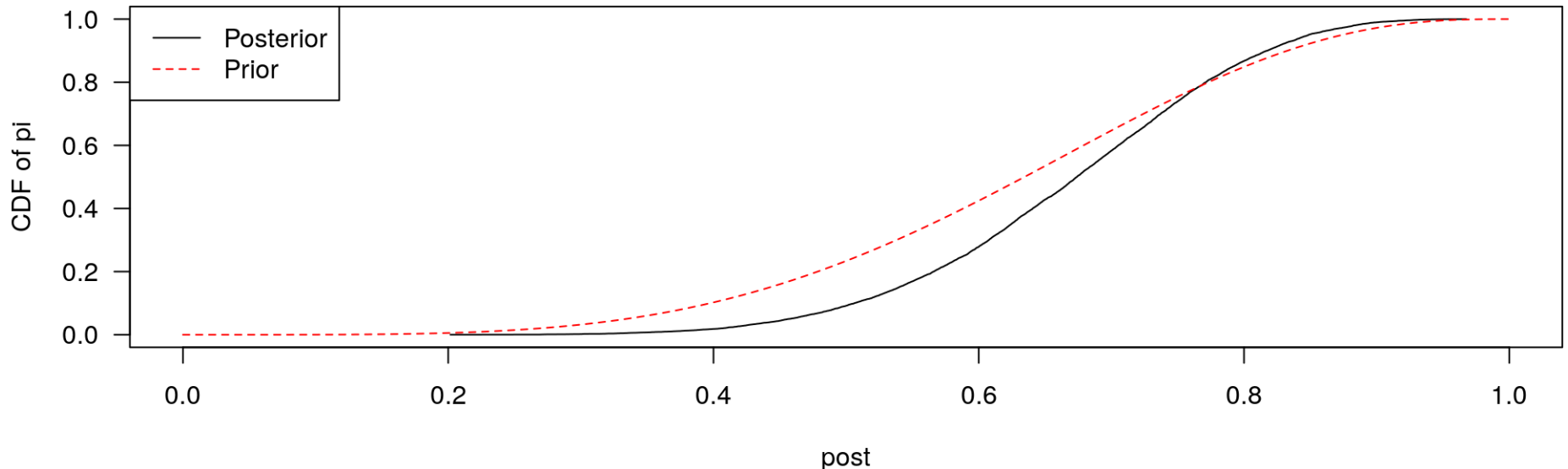
# Matching the Prior Predictive Distribution in Stan

```
functions { /* saved as ebola_rng.stan in R's working directory */
  vector ebola_rng(int S, real M, real m, int N, int y) { // must have a _rng suffix
    real common = 3 * (M - m);
    real a = (m * (4 * M - 3) + M) / common;
    real b = (m * (1 - 4 * M) + 5 * M - 2) / common;
    vector[S] post; // size equal to number of simulations
    int s = 1;
    if (a < 1 || b < 1) reject("M and m are inconsistent");
    while (s <= S) {
      real pi_tilde = beta_rng(a, b); // draw from prior distribution
      int y_tilde = binomial_rng(N, pi_tilde); // draw from prior predictive distribution
      if (y_tilde == y) { // keep iff matches observed data
        post[s] = pi_tilde; // indexing (mostly) works like in R
        s += 1; // s += 1; is equivalent to s = s + 1;
      }
    }
    return sort_asc(post); // easier to plot CDF if sorted ascending
  }
}
```

# Updating Beliefs about the Success Probability

```
rstan::expose_stan_functions("ebola_rng.stan") # puts Stan functions into R's workspace
M <- 2 / 3; m <- 0.635; a <- (m * (4 * M - 3) + M) / (3 * (M - m))
b <- (m * (1 - 4 * M) + 5 * M - 2) / (3 * (M - m))
N <- 7L; y <- 5L; S <- 10000L; post <- ebola_rng(S, M, m, N, y) # call as an R function
```

```
plot(post, y = (1:S) / S, type = "l", xlim = 0:1, ylab = "CDF of pi")
curve(pbeta(pi, a, b), from = 0, to = 1, add = TRUE, col = 2, lty = 2, xname = "pi")
legend("topleft", legend = c("Posterior", "Prior"), col = 1:2, lty = 1:2)
```



# Numerator of Bayes Rule

```
functions { /* saved as numerator.stan in R's working directory */
  vector numerator(vector pi, real M, real m, int N, int y) {
    real common = 3 * (M - m);
    real a = (m * (4 * M - 3) + M) / common;
    real b = (m * (1 - 4 * M) + 5 * M - 2) / common;
    int K = rows(pi); vector[K] log_pi = log(pi);
    vector[K] log1m_pi = log1m(pi); // log1m(x) == log(1 - x) but numerically accurate
    vector[K] log_prior = (a - 1) * log_pi + (b - 1) * log1m_pi - lbeta(a, b); // l prefix
    vector[K] log_likelihood = y * log_pi + (N - y) * log1m_pi + lchoose(N, y); // means log
    return exp(log_prior + log_likelihood); // equals exp(log_prior) * exp(log_likelihood)
  }
}
```

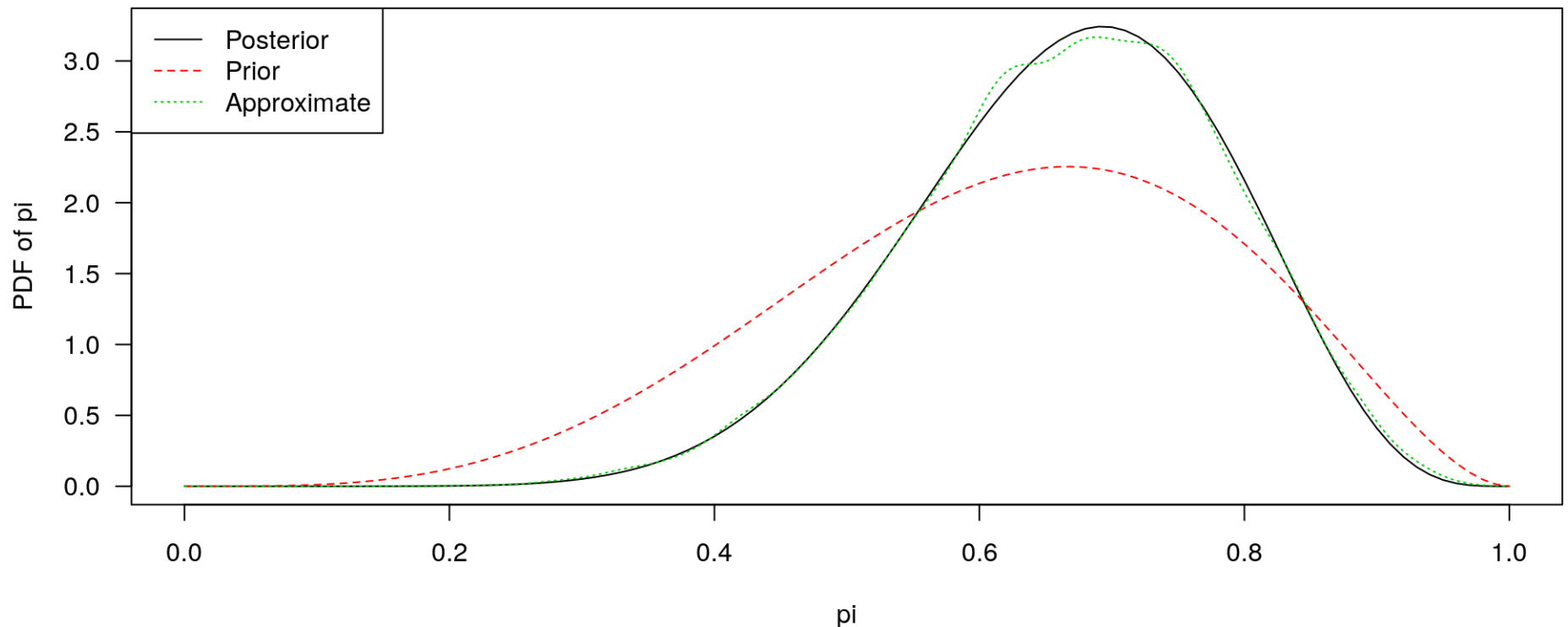
```
rstan::expose_stan_functions("numerator.stan")
```

```
(denom <- integrate(numerator, lower = 0, upper = 1, M = M, m = m, N = N, y = y)$value)
```

```
## [1] 0.2179323
```

# Same Info from the Probability Density Functions

```
curve(numerator(pi, M, m, N, y) / denom, from = 0, to = 1, xname = "pi", ylab = "PDF of pi")  
curve(dbeta(pi, a, b), from = 0, to = 1, col = 2, lty = 2, xname = "pi", add = TRUE)  
lines(density(post, from = 0, to = 1), col = 3, lty = 3)  
legend("topleft", legend = c("Posterior", "Prior", "Approximate"), col = 1:3, lty = 1:3)
```



# Deriving the Posterior Distribution Analytically

- Survivals are Binomial with probability  $\pi$ . Thus, BEFORE you see the results  $f(y|N, \pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$  while the Beta PDF is again  $f(\pi|a, b) = \frac{1}{B(a, b)} \pi^{a-1} (1 - \pi)^{b-1}$ , so  $B(a, b) = \int_0^1 \pi^{a-1} (1 - \pi)^{b-1} d\pi$
- AFTER  $y = 5$  out of  $N = 7$  people were cured, what are your beliefs about  $\pi$ ?

$$f(\pi|a, b, N, y) = \frac{f(\pi|a, b) L(\pi; N, y)}{\int_0^1 f(\pi|a, b) L(\pi; N, y) d\pi} \propto$$

$$\pi^{a-1} (1 - \pi)^{b-1} \pi^y (1 - \pi)^{N-y} = \pi^{a+y-1} (1 - \pi)^{b+N-y-1} = \pi^{a^*-1} (1 - \pi)^{b^*-1}$$

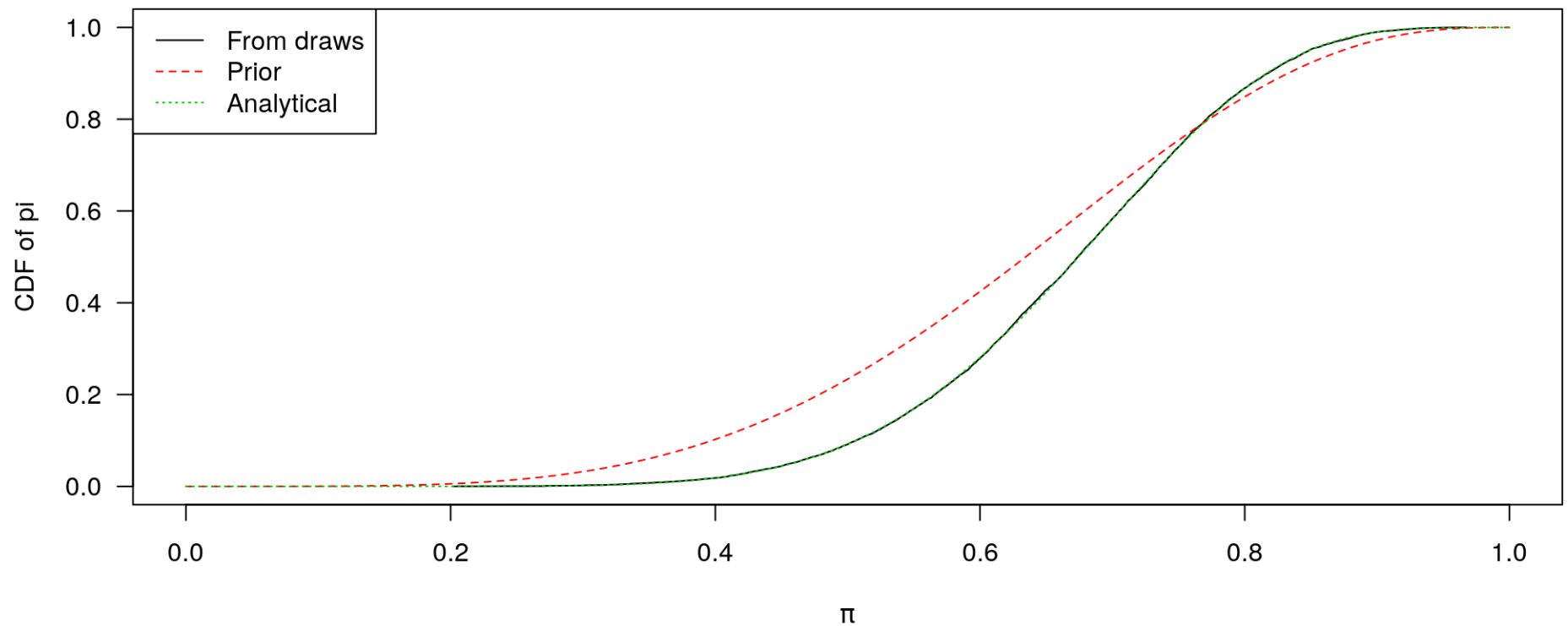
where  $a^* = a + y$  and  $b^* = b + N - y$

- $f(\pi|a^*, b^*)$  has the kernel of a Beta PDF and therefore the normalizing constant must be the reciprocal of  $B(a^*, b^*)$



# Checking the Posterior CDF

```
plot(post, y = (1:S) / S, type = "l", xlim = 0:1, xlab = expression(pi), ylab = "CDF of pi")  
curve(pbeta(pi, a, b), from = 0, to = 1, add = TRUE, col = 2, lty = 2, xname = "pi")  
curve(pbeta(pi, a + y, b + N - y), from = 0, to = 1, add = TRUE, col = 3, lty = 3, xname = "pi")  
legend("topleft", legend = c("From draws", "Prior", "Analytical"), col = 1:3, lty = 1:3)
```

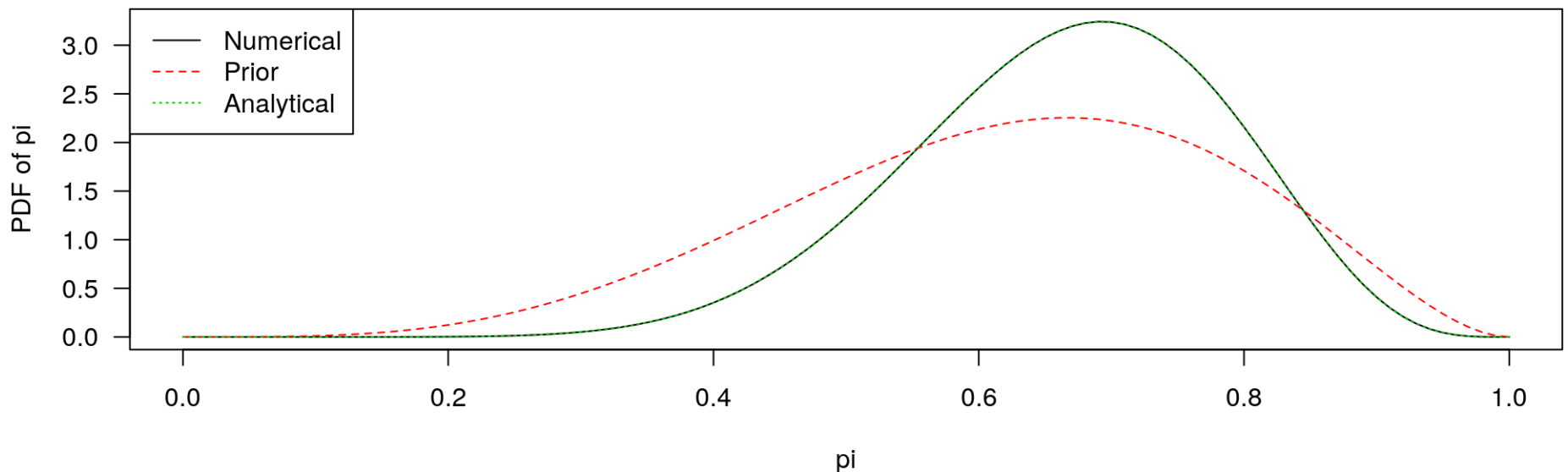


# Checking the Posterior PDF

```
c(exact = (a + y) / (a + y + b + N - y), approximate = mean(post)) # posterior expectation
```

```
##          exact approximate  
## 0.6666667 0.6665921
```

```
curve(numerator(pi, M, m, N, y) / denom, from = 0, to = 1, xname = "pi", ylab = "PDF of pi")  
curve(dbeta(pi, a, b), from = 0, to = 1, col = 2, lty = 2, xname = "pi", add = TRUE)  
curve(dbeta(pi, a + y, b + N - y), from = 0, to = 1, col = 3, lty = 3, xname = "pi", add = TRUE)  
legend("topleft", legend = c("Numerical", "Prior", "Analytical"), col = 1:3, lty = 1:3)
```



# Properties of this Posterior Distribution

- The first approach illustrates a key point: The posterior PDF is the function closest to the prior PDF that satisfies a constraint given by the observed data
- Posterior expectation is between the prior expectation and sample mean

```
c(prior = a / (a + b), posterior = (a + y) / (a + y + b + N - y), sample = y / N)
```

```
##      prior posterior      sample  
## 0.6232877 0.6666667 0.7142857
```

- As  $N \uparrow \infty$  with  $a$  and  $b$  fixed, the posterior mean approaches  $\frac{y}{N}$
- It does not matter if the data arrives one observation at a time, all at once, or somewhere in between. After  $N$  tries and  $y$  successes your posterior distribution will be the same, namely Beta with parameters  $a^* = a + y$  and  $b^* = b + N - y$ , and contains all the information available from past data
- Ergo, you can use your posterior PDF as your prior PDF for the next dataset