16: Beta

Amrita Kaur
July 27, 2022
Which video?

Let’s Explore

👍 10,000 👎 50

Others

👍 10 👎 0
Philosophical Ponderings

You ask two friends about the probability of rain tomorrow.

**Friend 1:** My leg itches when it rains and it's kind of itchy so, uh, $p=0.80$

**Friend 2:** I have done lots of complex calculations and have seen 10,451 days like tomorrow, so $p=0.80$

What’s the difference between these two estimates?
What if you don’t know a probability?
What if you don’t know a probability?
New definition of probability

Flip a frisbee $n + m$ times, get $n$ heads and $m$ tails. We don’t know the probability $p$ that the frisbee comes up heads.

**Frequentist**

$p$ is a single value.

$$p = \lim_{n+m \to \infty} \frac{n}{n + m} \approx \frac{n}{n + m}$$

**Bayesian**

$p$ is a random variable.
Flipping frisbees

Flip a frisbee 10 times, get 9 heads and 1 tails.
We don’t know the probability $p$ that the frisbee comes up heads.

**Frequentist**

\[ p = \frac{9}{10} \]

**Bayesian**

$p$ is a random variable.
Back to Bayes

Let $X$ be a RV representing the probability of heads

$$P(X = x|H = 9, T = 1) = \frac{P(H = 9, T = 1|X = x)f(X = x)}{P(H = 9, T = 1)}$$

$$= \frac{\binom{10}{9}x^9(1 - x)^1}{P(H = 9, T = 1)}$$

$$= K \cdot x^9(1 - x)^1$$
Graphically

\[ P(X = x | H = 9, T = 1) \]
Flipping, Generalized

Flip a frisbee $n + m$ times, get $n$ heads and $m$ tails.
- We don’t know the probability $X$ that the frisbee comes up heads.
- Belief before flipping is that $X \sim \text{Uni}(0,1)$
- Let $N =$ the number of heads
- Given $X = x$, the coin flips are independent

\[
f(X = x | N = n) = \frac{P(N = n | X = x) f(X = x)}{P(N = n)} = \frac{(n+m)^n x^n (1-x)^m}{P(N = n)} = \frac{(n+m)}{P(N=n)} x^n (1-x)^m
\]

\[
= \frac{1}{c} x^n (1-x)^m \text{ where } c = \int_0^1 x^n (1-x)^m \, dx
\]
Another Example

\[ f_X(x) = \frac{1}{c} x^1 (1 - x)^7 \]
Equivalently

If you start with a $X \sim \text{Uni}(0,1)$ prior:

- Let $a = \text{num "successes"} + 1$
- Let $b = \text{num "failures"} + 1$

Your new belief about the probability is:

$$f_X(x) = \frac{1}{c} x^{a-1} (1 - x)^{b-1}$$

where $c = \int_0^1 x^{a-1} (1 - x)^{b-1} dx$
## Beta random variable

\[ X \sim \text{Beta}(a, b) \]

- \( a > 0, b > 0 \)
- Support of \( X \): (0, 1)

**PDF**
\[
f(x) = \frac{1}{B(a, b)} x^{a-1} (1 - x)^{b-1}
\]

- \( B(a, b) = \int_0^1 x^{a-1} (1 - x)^{b-1} dx \), normalizing constant

**Expectation**
\[
E[X] = \frac{a}{a + b}
\]

**Variance**
\[
\text{Var}(X) = \frac{ab}{(a + b)^2 (a + b + 1)}
\]
Beta RV PDFs
What if Prior is Beta?

$X$ is a random variable for probability

If our prior belief about $X$ was beta:

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

What is our posterior belief about $X$ after observing $n$ heads and $m$ tails?

$$f(X = x|N = n) = ??$$
What if Prior is Beta?

\[
f(X = x | N = n) = \frac{P(N = n | X = x) f(X = x)}{P(N = n)}
\]

\[
= \frac{(n+m) x^n (1-x)^m f(X = x)}{P(N = n)}
\]

\[
= \frac{(n+m) x^n (1-x)^m \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}}{P(N = n)}
\]

\[
= \frac{(n+m) \frac{1}{B(a, b)}}{P(N=n)} x^n (1-x)^m x^{a-1} (1-x)^{b-1}
\]

\[
= K \cdot x^n (1-x)^m x^{a-1} (1-x)^{b-1}
\]

\[
X | N \sim \text{Beta}(n + a, m + b)
\]
A Beta Understanding

Beta is a **conjugate distribution** for Beta, meaning:

- Prior and posterior parametric forms are the same
- Practically, conjugate means easy update:
  Add number of “heads” and “tails” seen to Beta parameters.
A Beta Understanding

Beta is a **conjugate distribution** for Bernoulli, meaning:

- Prior and posterior parametric forms are the same
- Practically, conjugate means easy update: Add number of “heads” and “tails” seen to Beta parameters.

You can **invent a prior** to express how biased you believe frisbee is a priori:

- $\theta \sim \text{Beta}(a, b)$: pretend you’ve conducted $(a + b - 2)$ imaginary trials, where $(a - 1)$ trial produced a head and $(b - 1)$ produced a tail
- Choosing $\text{Beta}(1, 1) = \text{Uni}(0, 1)$ means you don’t hold any prior beliefs

<table>
<thead>
<tr>
<th>Prior</th>
<th>Beta($a = n_{imag} + 1, b = m_{imag} + 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>Observe $n$ successes and $m$ failures</td>
</tr>
<tr>
<td>Posterior</td>
<td>Beta($a + n, b + m$)</td>
</tr>
</tbody>
</table>
New definition of probability

Flip a frisbee $n + m$ times, get $n$ heads and $m$ tails.
We don’t know the probability $p$ that the frisbee comes up heads.

Frequentist

$p$ is a single value.

$$p = \frac{9}{10}$$

Bayesian

$p$ is a random variable.
Medicinal Beta

• Before being tested, a medicine is believed to “work” 80% of the time.
• The medicine is administered to 20 patients.
• It “works” for 14, “doesn’t work” for 6.

What is your new belief that the drug “works”?

**Frequentist**

Let $p$ be the probability your drug works.

\[
p \approx \frac{14}{20} = 0.7
\]

**Bayesian**
Medicinal Beta

- Before being tested, a medicine is believed to “work” 80% of the time.
- The medicine is administered to 20 patients.
- It “works” for 14, “doesn’t work” for 6.

What is your new belief that the drug “works”?

Frequentist

Let $p$ be the probability your drug works.

$p \approx \frac{14}{20} = 0.7$

Bayesian

Let $X$ be the probability your drug works.

$X$ is a random variable.
Medicinal Beta

• Before being tested, a medicine is believed to “work” 80% of the time.
• The medicine is administered to 20 patients.
• It “works” for 14, “doesn’t work” for 6.

What is your new belief that the drug “works”?

Bayesian: \( X \sim \text{Beta} \)

Prior:
- \( X \sim \text{Beta}(a = 81, b = 21) \)
- \( X \sim \text{Beta}(a = 9, b = 3) \)
- \( X \sim \text{Beta}(a = 5, b = 2) \)

Interpretation:
- 80 successes/100 trials
- 8 successes/10 trials
- 4 successes/5 trials
Medicinal Beta

• Before being tested, a medicine is believed to “work” 80% of the time.
• The medicine is administered to 20 patients.
• It “works” for 14, “doesn’t work” for 6.

What is your new belief that the drug “works”?

Bayesian: \( X \sim \text{Beta} \)

Prior: \( X \sim \text{Beta}(a = 5, b = 2) \)

Posterior: \( X \sim \text{Beta}(a = 5 + 14, b = 2 + 6) \)
\( \sim \text{Beta}(a = 19, b = 8) \)

\[
E[X] = \frac{a}{a + b} = \frac{19}{19 + 8} \approx 0.70
\]

\[
\text{mode}(X) = \frac{a - 1}{a + b - 2} = \frac{18}{18 + 7} \approx 0.72
\]
Laplace Smoothing

Prior: \( X \sim \text{Beta}(a = 2, b = 2) \)

- One imaginary success
- One imaginary failure
Which video?

Let’s Explore

👍 10,000  👎 50

Others

👍 10  👎 0
Which video?

10,000

50

10

0
iid Random Variables
iid random variables

Consider $n$ variables $X_1, X_2, ..., X_n$.

$X_1, X_2, ..., X_n$ are **independent and identically distributed** if

- $X_1, X_2, ..., X_n$ are independent, and
- All have the same PMF (if discrete) or PDF (if continuous).

$\Rightarrow E[X_i] = \mu$ for $i = 1, ..., n$

$\Rightarrow \text{Var}(X_i) = \sigma^2$ for $i = 1, ..., n$

Notations: iid, i.i.d, IID
Quick check

Are \( X_1, X_2, \ldots, X_n \) i.i.d. with the following distributions?

1. \( X_i \sim \text{Exp}(\lambda), \ X_i \) independent

2. \( X_i \sim \text{Exp}(\lambda_i), \ X_i \) independent

3. \( X_i \sim \text{Exp}(\lambda), \ X_1 = X_2 = \cdots = X_n \)

4. \( X_i \sim \text{Bin}(n_i, p), \ X_i \) independent
Adding Random Variables
ELO ratings

Basketball == Stats

Skill  Magic  Determination

What is the probability that the Warriors win?
How do you model zero-sum games?
ELO ratings

Each team has an ELO score $S$, calculated based on its past performance.

- Each game, a team has ability $A \sim \mathcal{N}(S, 200^2)$.
- The team with the higher sampled ability wins.

What is the probability that Warriors win this game?

Want:

$P(\text{Warriors win}) = P(A_W > A_O)$
Probability of Winning

\[ P(\text{Warriors win}) = P(A_W > A_O) \]

\[ A_W \sim \mathcal{N}(S = 1657, 200^2) \]

\[ A_O \sim \mathcal{N}(S = 1470, 200^2) \]
Sum of Dice Rolls

Roll $n$ independent dice. Let $X_i$ be the outcome of roll $i$. $X_i$ are i.i.d.

$Y = \sum_{i=1}^{1} X_i$ \hspace{1cm} Sum of 1 die roll

$Y = \sum_{i=1}^{2} X_i$ \hspace{1cm} Sum of 2 dice rolls

$Y = \sum_{i=1}^{3} X_i$ \hspace{1cm} Sum of 3 dice rolls
Insight to Convolution

Imagine a game where each player independently scores between 0 and 100 points:

• Let $X$ be the amount of points you score.
• Let $Y$ be the amount of points your opponent scores.
• Let’s say you know $P(X = x)$ and $P(Y = y)$.

What is the probability of a tie?

$$P(\text{tie}) = \sum_{i=0}^{100} P(X = i, Y = i)$$

$$= \sum_{i=0}^{100} P(X = i)P(Y = i)$$