Beta Prior for Binomial

User

Prove me that the prior of a Binomial Distribution is Beta

Assistant chatgpt-4o-latest

2025-03-04, 09:28:07

To understand why the **Beta distribution** is the **conjugate prior** of the **Binomial distribution**, let's go through the derivation step by step. ### 1. Problem Setup We assume a **Bernoulli/Binomial likelihood** and a **Beta prior**: - **Likelihood (Binomial Model)**: Let \(X \sim \text{Bin}(n, \theta) \), meaning \[P(X = k \mid \theta) = \binom{n}{k} \theta^k (1 - \theta)^{n - k} \] for \(k = 0, 1, \dots, n \). - **Prior (Beta Distribution)**: Assume that \(\theta \) follows a **Beta prior**: \[P(\theta) = \frac{\theta^{\alpha - 1} (1 - \theta)^{\beta - 1}}{B(\alpha, \beta)} \] where \(B(\alpha, \beta) \) is the Beta function ensuring normalization. We want to find the **posterior** distribution \(P(\theta \mid k) \). ### 2. Compute the Posterior Using Bayes' theorem: \[P(\theta \mid X = k) \propto P(X = k \mid \theta) P(\theta) \] Substituting the Binomial likelihood and the Beta prior: \[P(\theta \mid X = k) \propto \left[\binom{n}{k} \theta^k (1 - \theta)^{n-k} \right] \times \left[\frac{\theta^{\alpha - 1} (1 - \theta)^{\beta - 1}}{B(\alpha, \beta)} \right] \] Ignoring constants that do not depend on \(\theta \) (such as the binomial coefficient and Beta function), we get: \[P(\theta \mid X = k) \propto \theta^k (1 - \theta)^{n-k} \times \theta^{\alpha - 1} (1 - \theta)^{\beta - 1} \] which simplifies to: \[P(\theta \mid X = k) \propto \theta^{(\alpha + k) - 1} (1 - \theta)^{(\beta + n - k) - 1} \] ### 3. Recognizing the Result as a Beta Distribution Comparing with the form of a Beta distribution: \[\text{Beta}(\alpha', \beta'), \quad \text{where} \quad \alpha' = \alpha + k, \quad \beta' = \beta + n - k. \] Thus, the posterior follows: \[P(\theta \mid X = k) = \text{Beta}(\alpha + k, \beta + n - k). \] ### Conclusion Since the posterior distribution is also a Beta distribution, we have shown that the **Beta distribution is the conjugate prior of the Binomial distribution**. This property is useful in Bayesian inference, as it ensures the posterior remains in the same family, simplifying analysis and computation.

User

2025-03-04, 09:30:52

But why did we assume a Beta prior for the Binomial Distribution?

Assistant chatgpt-4o-latest

2025-03-04, 09:30:52

Great question! The reason we assume a **Beta prior** for a **Binomial distribution** is based on both theoretical and practical considerations. Let's break it down. --- ### 1. **The Need for a Prior Distribution** In Bayesian inference, we treat the probability parameter \(\theta \) of a Binomial process as an **uncertain quantity**, meaning that before we observe any data, we describe \(\theta \) using a **prior distribution** \(P(\theta) \). This prior reflects our beliefs about \(\theta \) before collecting data. But what should we choose as \(P(\theta) \)? --- ### 2. **Criteria for Choosing a Prior** A good prior distribution should: - **Be defined on the same domain as \(\theta \)**: Since \(\theta \) is a probability, it must lie in \([0,1] \). Thus, we need a prior distribution that is **supported on \([0,1] \)**. - **Be mathematically convenient**: The posterior distribution \(P(\theta \mid X) \) should be easy to compute. - **Reflect prior knowledge (if available)**: If we believe \(\theta \) is likely around a particular value before observing data, our prior should capture that belief. --- ### 3. **Beta Distribution Satisfies These Criteria** The **Beta distribution** is particularly well-suited as a prior for a Binomial proportion \(\theta \) because: 1. **Its domain is [0,1]**, matching the range of \(\theta \). 2. **It has a flexible shape**: The Beta distribution has two parameters \((\alpha, \beta) \) that allow it to represent different beliefs about \(\theta \), such as: - **Uniform prior**: \(\text{Beta}(1,1) \) is the uniform distribution, meaning we have no preference for any particular \(\theta \). - **Informative prior**: Different choices of \((\alpha, \beta) \) can represent prior beliefs about the likelihood of success \(\theta \), such as being skewed towards higher or lower values. 3. **It is the conjugate prior of the Binomial**: - When we assume a Beta prior, the posterior distribution after observing data remains **Beta**, just with updated parameters \((\alpha + k, \beta + n - k) \). - This makes Bayesian updating simple and computationally efficient, since the posterior is in the same family as the prior. --- ### 4. **Derivation of Why Beta is Conjugate to Binomial** We saw in the previous response that, given: - A **Binomial likelihood**: \[P(X = k \mid \theta) = \binom{n}{k} \theta^k (1 - \theta)^{n - k} \] - A **Beta prior**: \[P(\theta) = \frac{\theta^{\alpha - 1} (1 - \theta)^{\beta - 1}}{B(\alpha, \beta)} \] Then, after applying Bayes’ theorem, the posterior remains a **Beta distribution**: \[P(\theta \mid X = k) = \text{Beta}(\alpha + k, \beta + n - k). \] This conjugacy property makes Beta a **natural choice** for modeling uncertainty in \(\theta \). --- ### 5. **What If We Chose a Different Prior?** If we chose a non-Beta prior (e.g., a uniform distribution on some other domain or a normal distribution), the posterior would not have a simple closed form. Instead, we'd have to: - Calculate the posterior using **numerical integration** or approximations. - Lose the convenient **Bayesian updating rule**. Thus, choosing a Beta prior simplifies both the theory and computation. --- ### **Final Answer: Why Assume a Beta Prior for a Binomial?** We assume a **Beta prior** for the Binomial distribution because: 1. **The domain matches**: Both \(\theta \) and the Beta distribution are defined on \([0,1] \). 2. **Beta is flexible**: It can represent different prior beliefs about \(\theta \). 3. **It is conjugate**: The posterior remains Beta, which simplifies Bayesian updating. 4. **It allows for efficient computations**: We can update beliefs about \(\theta \) easily after observing data. This is why the **Beta distribution is the natural prior for a Binomial model** in Bayesian inference.