

11.7: The Beta-Bernoulli Process

An interesting thing to do in almost any parametric probability model is to “randomize” one or more of the parameters. Done in a clever way, this often leads to interesting new models and unexpected connections between models. In this section we will randomize the success parameter in the Bernoulli trials process. This leads to interesting and surprising connections with Pólya's urn process.

Basic Theory

Definitions

First, recall that the beta distribution with left parameter $a \in (0, \infty)$ and right parameter $b \in (0, \infty)$ is a continuous distribution on the interval $(0, 1)$ with probability density function g given by

$$g(p) = \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1}, \quad p \in (0, 1) \quad (11.7.1)$$

where B is the *beta function*. So $B(a, b)$ is simply the normalizing constant for the function $p \mapsto p^{a-1} (1-p)^{b-1}$ on the interval $(0, 1)$. Here is our main definition:

Suppose that P has the beta distribution with left parameter $a \in (0, \infty)$ and right parameter $b \in (0, \infty)$. Next suppose that $\mathbf{X} = (X_1, X_2, \dots)$ is a sequence of indicator random variables with the property that given $P = p \in (0, 1)$, \mathbf{X} is a conditionally independent sequence with

$$\mathbb{P}(X_i = 1 \mid P = p) = p, \quad i \in \mathbb{N}_+ \quad (11.7.2)$$

Then \mathbf{X} is the *beta-Bernoulli process* with parameters a and b .

In short, given $P = p$, the sequence \mathbf{X} is a Bernoulli trials sequence with success parameter p . In the usual language of reliability, X_i is the outcome of trial i , where 1 denotes success and 0 denotes failure. For a specific application, suppose that we select a random probability of heads according to the beta distribution with parameters a and b , and then toss a coin with this probability of heads repeatedly.

Outcome Variables

What's our first step? Well, of course we need to compute the finite dimensional distributions of \mathbf{X} . Recall that for $r \in \mathbb{R}$ and $j \in \mathbb{N}$, $r^{[j]}$ denotes the ascending power $r(r+1) \cdots [r+(j-1)]$. By convention, a product over an empty index set is 1, so $r^{[0]} = 1$.

Suppose that $n \in \mathbb{N}_+$ and $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$. Let $k = x_1 + x_2 + \cdots + x_n$. Then

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}} \quad (11.7.3)$$

Proof

First, note that $\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid P = p) = p^k (1-p)^{n-k}$ by the conditional independence. Thus, conditioning on P gives

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mathbb{E}[\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n \mid P)] \quad (11.7.4)$$

$$= \int_0^1 p^k (1-p)^{n-k} \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} dp \quad (11.7.5)$$

$$= \frac{B[a+k, b+(n-k)]}{B(a, b)} = \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}} \quad (11.7.6)$$

The last step uses a property of the beta function.

From this result, it follows that Pólya's urn process with parameters $a, b, c \in \mathbb{N}_+$ is equivalent to the beta-Bernoulli process with parameters a/c and b/c , quite an interesting result. Note that since the joint distribution above depends only on

$x_1 + x_2 + \dots + x_n$, the sequence \mathbf{X} is exchangeable. Finally, it's interesting to note that the beta-Bernoulli process with parameters a and b could simply be defined as the sequence with the finite-dimensional distributions above, without reference to the beta distribution! It turns out that *every* exchangeable sequence of indicator random variables can be obtained by randomizing the success parameter in a sequence of Bernoulli trials. This is *de Finetti's theorem*, named for Bruno de Finetti, which is studied in the section on backwards martingales.

For each $i \in \mathbb{N}_+$

1. $\mathbb{E}(X_i) = \frac{a}{a+b}$
2. $\text{var}(X_i) = \frac{a}{a+b} \frac{b}{a+b}$

Proof

Since the sequence is exchangeable, X_i has the same distribution as X_1 , so $\mathbb{P}(X_i = 1) = \frac{a}{a+b}$. The mean and variance now follow from standard results for indicator variables.

Thus \mathbf{X} is a sequence of identically distributed variables, quite surprising at first but of course inevitable for any exchangeable sequence. Compare the [joint distribution](#) with the [marginal distributions](#). Clearly the variables are dependent, so let's compute the covariance and correlation of a pair of outcome variables.

Suppose that $i, j \in \mathbb{N}_+$ are distinct. Then

1. $\text{cov}(X_i, X_j) = \frac{ab}{(a+b)^2(a+b+1)}$
2. $\text{cor}(X_i, X_j) = \frac{1}{a+b+1}$

Proof

Since the variables are exchangeable, $\mathbb{P}(X_i = 1, X_j = 1) = \mathbb{P}(X_1 = 1, X_2 = 1) = \frac{a}{a+b} \frac{a+1}{a+b+1}$. The results now follow from standard formulas for covariance and correlation.

Thus, the variables are positively correlated. It turns out that in any *infinite* sequence of exchangeable variables, the the variables must be nonnegatively correlated. Here is another result that explores how the variables are related.

Suppose that $n \in \mathbb{N}_+$ and $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$. Let $k = \sum_{i=1}^n x_i$. Then

$$\mathbb{P}(X_{n+1} = 1 \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{a+k}{a+b+n} \quad (11.7.7)$$

Proof

Using the [joint distribution](#),

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1 \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) &= \frac{\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n, X_{n+1} = 1)}{\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)} \\ &= \frac{a^{[k+1]} b^{[n-k]} (a+b)^{[n]}}{(a+b)^{[n+1]} a^{[k]} b^{[n-k]}} = \frac{a+k}{a+b+n} \end{aligned}$$

The beta-Bernoulli model starts with the conditional distribution of \mathbf{X} given P . Let's find the conditional distribution in the other direction.

Suppose that $n \in \mathbb{N}_+$ and $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$. Let $k = \sum_{i=1}^n x_i$. Then the conditional distribution of P given $(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ is beta with left parameter $a+k$ and right parameter $b+(n-k)$. Hence

$$\mathbb{E}(P \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{a+k}{a+b+k} \quad (11.7.8)$$

Proof

This follows from Bayes' theorem. The conditional PDF $g(\cdot \mid x_1, x_2, \dots, x_n)$ is given by

$$g(p | X_1, x_2, \dots, x_n) = \frac{g(p)\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | P = p)}{\int_0^1 g(t)\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | P = t)dt}, \quad p \in (0, 1) \quad (11.7.9)$$

The numerator is

$$\frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} p^k (1-p)^{n-k} = \frac{1}{B(a, b)} p^{a+k-1} (1-p)^{b+n-k-1} \quad (11.7.10)$$

The denominator is simply the normalizing constant for the expression, as a function of p and is $B(a+k, b+n-k)/B(a, b)$. Hence

$$g(p | k) = \frac{1}{B(a+k, b+n-k)} p^{a+k-1} (1-p)^{b+n-k-1}, \quad p \in (0, 1) \quad (11.7.11)$$

The last result follows since the mean of the beta distribution is the left parameter divided by the sum of the parameters.

Thus, the left parameter increases by the number of successes while the right parameter increases by the number of failures. In the language of Bayesian statistics, the original distribution of P is the *prior distribution*, and the conditional distribution of P given the data (x_1, x_2, \dots, x_n) is the *posterior distribution*. The fact that the posterior distribution is beta whenever the prior distribution is beta means that the beta distributions is *conjugate* to the Bernoulli distribution. The conditional expected value in the last theorem is the *Bayesian estimate* of p when p is modeled by the random variable P . These concepts are studied in more generality in the section on Bayes Estimators in the chapter on Point Estimation. It's also interesting to note that the expected values in the last two theorems are the same: If $n \in \mathbb{N}$, $(x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ and $k = \sum_{i=1}^n x_i$ then

$$\mathbb{E}(X_{n+1} | X_1 = x_1, \dots, X_n = x_n) = \mathbb{E}(P | X_1 = x_1, \dots, X_n = x_n) = \frac{a+k}{a+b+n} \quad (11.7.12)$$

Run the simulation of the beta coin experiment for various values of the parameter. Note how the posterior probability density function changes from the prior probability density function, given the number of heads.

The Number of Successes

It's already clear that the number of successes in a given number of trials plays an important role, so let's study these variables. For $n \in \mathbb{N}_+$, let

$$Y_n = \sum_{i=1}^n X_i \quad (11.7.13)$$

denote the number of successes in the first n trials. Of course, $\mathbf{Y} = (Y_0, Y_1, \dots)$ is the partial sum process associated with $\mathbf{X} = (X_1, X_2, \dots)$.

Y_n has probability density function given by

$$\mathbb{P}(Y_n = k) = \binom{n}{k} \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}}, \quad k \in \{0, 1, \dots, n\} \quad (11.7.14)$$

Proof

Every bit string of length n with 1 occurring exactly k times has the probability given in the [joint distribution](#) above. There are $\binom{n}{k}$ such bit strings.

The distribution of Y_n is known as the *beta-binomial distribution* with parameters n , a , and b .

In the simulation of the beta-binomial experiment, vary the parameters and note how the shape of the probability density function of Y_n (discrete) parallels the shape of the probability density function of P (continuous). For various values of the parameters, run the simulation 1000 times and compare the empirical density function to the probability density function.

The case where the parameters are both 1 is interesting.

If $a = b = 1$, so that P is uniformly distributed on $(0, 1)$, then Y_n is uniformly distributed on $\{0, 1, \dots, n\}$.

Proof

Note that $1^{[j]} = j!$ and $2^{[j]} = (j+1)!$ for $j \in \mathbb{N}$. Hence, from the [general PDF](#) Y_n above

$$\mathbb{P}(Y_n = k) = \frac{n!}{k!(n-k)!} \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}, \quad k \in \{0, 1, \dots, n\} \quad (11.7.15)$$

Next, let's compute the mean and variance of Y_n .

The mean and variance of Y_n are

1. $\mathbb{E}(Y_n) = n \frac{a}{a+b}$
2. $\text{var}(Y_n) = n \frac{ab}{(a+b)^2} \left[1 + (n-1) \frac{1}{a+b+1} \right]$

Proof

These results follow from the [mean](#) and [covariance](#) results given above:

$$\mathbb{E}(Y_n) = \sum_{i=a}^n \mathbb{E}(X_i) = n \frac{a}{a+b} \quad (11.7.16)$$

$$\text{var}(Y_n) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, X_j) = n \frac{ab}{(a+b)^2} + n(n-1) \frac{ab}{(a+b)^2(a+b+1)} \quad (11.7.17)$$

In the simulation of the beta-binomial experiment, vary the parameters and note the location and size of the mean-standard deviation bar. For various values of the parameters, run the simulation 1000 times and compare the empirical moments to the true moments.

We can restate the conditional distributions in the last subsection more elegantly in terms of Y_n .

Let $n \in \mathbb{N}$.

1. The conditional distribution of X_{n+1} given Y_n is

$$\mathbb{P}(X_{n+1} = 1 \mid Y_n) = \mathbb{E}(X_{n+1} \mid Y_n) = \frac{a + Y_n}{a + b + n} \quad (11.7.18)$$

2. The conditional distribution of P given Y_n is beta with left parameter $a + Y_n$ and right parameter $b + (n - Y_n)$. In particular

$$\mathbb{E}(P \mid Y_n) = \frac{a + Y_n}{a + b + n} \quad (11.7.19)$$

Proof

The proof is easy using the nesting property of conditional expected value and the fact that the conditional distributions given (X_1, X_2, \dots, X_n) depend only on $Y_n = \sum_{i=1}^n X_i$.

1. Note that

$$\mathbb{E}(X_{n+1} \mid Y_n) = \mathbb{E}[\mathbb{E}(X_{n+1} \mid Y_n) \mid X_1, X_2, \dots, X_n] \quad (11.7.20)$$

$$= \mathbb{E}\left[\frac{a + Y_n}{a + b + n} \mid X_1, X_2, \dots, X_n\right] = \mathbb{E}\left(\frac{a + Y_n}{a + b + n} \mid Y_n\right) = \frac{a + Y_n}{a + b + n} \quad (11.7.21)$$

2. Similarly, if $A \subseteq (0, 1)$ is measurable then $\mathbb{P}(P \in A \mid X_1, X_2, \dots, X_n)$ depends only on Y_n and so

$$\mathbb{P}(P \in A \mid Y_n) = \mathbb{E}[\mathbb{P}(P \in A \mid Y_n) \mid X_1, X_2, \dots, X_n] \quad (11.7.22)$$

$$= \mathbb{E}[\mathbb{P}(P \in A \mid X_1, X_2, \dots, X_n) \mid Y_n] = \mathbb{P}(P \in A \mid Y_n) \quad (11.7.23)$$

Once again, the conditional expected value $\mathbb{E}(P \mid Y_n)$ is the *Bayesian estimator* of p . In particular, if $a = b = 1$, so that P has the uniform distribution on $(0, 1)$, then $\mathbb{P}(X_{n+1} = 1 \mid Y_n = n) = \frac{n+1}{n+2}$. This is *Laplace's rule of succession*, another interesting

connection. The rule is named for Pierre Simon Laplace, and is studied from a different point of view in the section on Independence.

The Proportion of Successes

For $n \in \mathbb{N}_+$, let

$$M_n = \frac{Y_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \quad (11.7.24)$$

so that M_n is the *sample mean* of (X_1, X_2, \dots, X_n) , or equivalently the proportion of successes in the first n trials. Properties of M_n follow easily from the corresponding properties of Y_n . In particular, $\mathbb{P}(M_n = k/n) = \mathbb{P}(Y_n = k)$ for $k \in \{0, 1, \dots, n\}$ as given [above](#), so let's move on to the mean and variance.

For $n \in \mathbb{N}_+$, the mean and variance of M_n are

1. $\mathbb{E}(M_n) = \frac{a}{a+b}$
2. $\text{var}(M_n) = \frac{1}{n} \frac{ab}{(a+b)^2} + \frac{n-1}{n} \frac{ab}{(a+b)^2(a+b+1)}$

Proof

These results follow from the mean and variance of Y_n [above](#) and properties of expected value and variance:

1. $\mathbb{E}(M_n) = \frac{1}{n} \mathbb{E}(Y_n)$
2. $\text{var}(M_n) = \frac{1}{n^2} \text{var}(Y_n)$

So $\mathbb{E}(M_n)$ is constant in $n \in \mathbb{N}_+$ while $\text{var}(M_n) \rightarrow ab/(a+b)^2(a+b+1)$ as $n \rightarrow \infty$. These results suggest that perhaps M_n has a limit, in some sense, as $n \rightarrow \infty$. For an ordinary sequence of Bernoulli trials with success parameter $p \in (0, 1)$, we know from the law of large numbers that $M_n \rightarrow p$ as $n \rightarrow \infty$ with probability 1 and in mean (and hence also in distribution). What happens here when the success probability P has been randomized with the beta distribution? The answer is what we might hope.

$M_n \rightarrow P$ as $n \rightarrow \infty$ with probability 1 and in mean square, and hence also in in distribution.

Proof

Let g denote the PDF of P . For convergence with probability 1, we condition on P

$$\mathbb{P}(M_n \rightarrow P \text{ as } n \rightarrow \infty) = \mathbb{E}[\mathbb{P}(M_n \rightarrow P \text{ as } n \rightarrow \infty) \mid P] \quad (11.7.25)$$

$$= \int_0^1 \mathbb{P}(M_n \rightarrow p \text{ as } n \rightarrow \infty \mid P = p) g(p) dp = \int_0^1 g(p) dp = 1 \quad (11.7.26)$$

For convergence in mean square, once again we condition on P . Note that

$$\mathbb{E}[(M_n - P)^2 \mid P = p] = \mathbb{E}[(M_n - p)^2 \mid P = p] = \frac{p(1-p)}{n} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (11.7.27)$$

Hence by the dominated convergence theorem,

$$\mathbb{E}[(M_n - P)^2] = \int_0^1 \frac{p(1-p)}{n} g(p) dp \rightarrow 0 \text{ as } n \rightarrow \infty \quad (11.7.28)$$

Proof of convergence in distribution

Convergence with probability 1 implies convergence in distribution, but it's interesting to give a direct proof. For $x \in (0, 1)$, note that

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(Y_n \leq nx) = \sum_{k=0}^{\lfloor nx \rfloor} \binom{n}{k} \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}} \quad (11.7.29)$$

where $\lfloor \cdot \rfloor$ is the floor function. But recall that

$$\frac{a^{[k]}b^{[n-k]}}{(a+b)^{[n]}} = \frac{B(a+k, b+n-k)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^1 p^{a+k-1} (1-p)^{b+n-k-1} dp \quad (11.7.30)$$

Substituting and doing some algebra we get

$$\mathbb{P}(M_n \leq x) = \frac{1}{B(a, b)} \int_0^1 \left[\sum_{k=0}^{\lfloor nx \rfloor} \binom{n}{k} p^k (1-p)^{n-k} \right] p^{a-1} (1-p)^{b-1} dp \quad (11.7.31)$$

The sum in the square brackets is $\mathbb{P}(W_n \leq nx) = \mathbb{P}(W_n/n \leq x)$ where W_n has the ordinary binomial distribution with parameters n and p . But W_n/n converges (in every sense) to p as $n \rightarrow \infty$ so $\mathbb{P}(W_n/n \leq x) \rightarrow \mathbf{1}(p \leq x)$ as $n \rightarrow \infty$. So by the dominated convergence theorem,

$$\mathbb{P}(M_n \leq x) \rightarrow \frac{1}{B(a, b)} \int_0^x p^{a-1} (1-p)^{b-1} dp = \mathbb{P}(P \leq x) \quad (11.7.32)$$

Recall again that the Bayesian estimator of p based on (X_1, X_2, \dots, X_n) is

$$\mathbb{E}(P | Y_n) = \frac{a + Y_n}{a + b + n} = \frac{a/n + M_n}{a/n + b/n + 1} \quad (11.7.33)$$

It follows from the last theorem that $\mathbb{E}(P | Y_n) \rightarrow P$ with probability 1, in mean square, and in distribution. The stochastic process $\mathbf{Z} = \{Z_n = (a + Y_n)/(a + b + n) : n \in \mathbb{N}\}$ that we have seen several times now is of fundamental importance, and turns out to be a martingale. The theory of martingales provides powerful tools for studying convergence in the beta-Bernoulli process.

The Trial Number of a Success

For $k \in \mathbb{N}_+$, let V_k denote the trial number of the k th success. As we have seen before in similar circumstances, the process $\mathbf{V} = (V_1, V_2, \dots)$ can be defined in terms of the process \mathbf{Y} :

$$V_k = \min\{n \in \mathbb{N}_+ : Y_n = k\}, \quad k \in \mathbb{N}_+ \quad (11.7.34)$$

Note that V_k takes values in $\{k, k+1, \dots\}$. The random processes $\mathbf{V} = (V_1, V_2, \dots)$ and $\mathbf{Y} = (Y_0, Y_1, \dots)$ are inverses of each other in a sense.

For $k \in \mathbb{N}$ and $n \in \mathbb{N}_+$ with $k \leq n$,

1. $V_k \leq n$ if and only if $Y_n \geq k$
2. $V_k = n$ if and only if $Y_{n-1} = k-1$ and $X_n = 1$

The probability density function of V_k is given by

$$\mathbb{P}(V_k = n) = \binom{n-1}{k-1} \frac{a^{[k]}b^{[n-k]}}{(a+b)^{[n]}}, \quad n \in \{k, k+1, \dots\} \quad (11.7.35)$$

Proof 1

As usual, we can condition on P and use known results for ordinary Bernoulli trials. Given $P = p$, random variable V_k has the negative binomial distribution with parameters k and p . Hence

$$\begin{aligned} \mathbb{P}(V_k = n) &= \int_0^1 \mathbb{P}(V_k = n | P = p) g(p) dp = \int_0^1 \binom{n-1}{k-1} p^k (1-p)^{n-k} \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1} dp \\ &= \binom{n-1}{k-1} \frac{1}{B(a, b)} \int_0^1 p^{a+k-1} (1-p)^{b+n-k-1} dp \\ &= \binom{n-1}{k-1} \frac{B(a+k, b+n-k)}{B(a, b)} = \binom{n-1}{k-1} \frac{a^{[k]}b^{[n-k]}}{(a+b)^{[n]}} \end{aligned}$$

Proof 2

In this proof, we condition on Y_{n-1} . Using the PDF of Y_{n-1} and the result [above](#),

$$\begin{aligned}\mathbb{P}(V_k = n) &= \mathbb{P}(Y_{n-1} = k-1, X_n = 1) = \mathbb{P}(Y_{n-1} = k-1) \mathbb{P}(X_n = 1 \mid Y_{n-1} = k-1) \\ &= \binom{n-1}{k-1} \frac{a^{[k-1]} b^{[(n-1)-(k-1)]}}{(a+b)^{[n-1]}} \frac{a+k-1}{a+b+(n-1)} = \binom{n-1}{k-1} \frac{a^{[k]} b^{[n-k]}}{(a+b)^{[n]}}\end{aligned}$$

The distribution of V_k is known as the *beta-negative binomial distribution* with parameters k , a , and b .

If $a = b = 1$ so that P is uniformly distributed on $(0, 1)$, then

$$\mathbb{P}(V_k = n) = \frac{k}{n(n+1)}, \quad n \in \{k, k+1, k+2, \dots\} \quad (11.7.36)$$

Proof

Recall again that $1^{[j]} = j!$ and $2^{[j]} = (j+1)!$ for $j \in \mathbb{N}$. Hence from the previous result,

$$\mathbb{P}(V_k = n) = \frac{(n-1)!}{(k-1)!(n-k)!} \frac{k!(n-k)!}{(n+1)!} = \frac{k}{n(n+1)}, \quad n \in \{k, k+1, \dots\} \quad (11.7.37)$$

In the simulation of the beta-negative binomial experiment, vary the parameters and note the shape of the probability density function. For various values of the parameters, run the simulation 1000 times and compare the empirical density function to the probability density function.

The mean and variance of V_k are

1. $\mathbb{E}(V_k) = k \frac{a+b-1}{a-1}$ if $a > 1$.
2. $\text{var}(V_k) = k \frac{a+b-1}{(a-1)(a-2)} [b + k(a+b-2)] - k^2 \left(\frac{a+b-1}{a-1} \right)^2$

Proof

From our work with the negative binomial distribution we know that $\mathbb{E}(V_k | P = p) = k \frac{1}{p}$ and $\mathbb{E}(V_k^2 | P = p) = k \frac{1-p}{p^2} + \frac{k^2}{p^2}$.

Thus, conditioning on P we have

$$\mathbb{E}(V_k) = \mathbb{E}[\mathbb{E}(V_k | P)] = \int_0^1 \frac{k}{p} \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)} = k \frac{B(a-1, b)}{B(a, b)} = k \frac{a+b-1}{a-1} \quad (11.7.38)$$

which gives part (a). Similarly

$$\mathbb{E}(V_k^2) = \mathbb{E}[\mathbb{E}(V_k^2 | P)] = \int_0^1 \left(k \frac{1-p}{p^2} + \frac{k^2}{p^2} \right) \frac{p^{a-1}(1-p)^{b-1}}{B(a, b)} \quad (11.7.39)$$

$$= k \frac{B(a-2, b+1)}{B(a, b)} + k^2 \frac{B(a-2, b)}{B(a, b)} = k \frac{b(a+b-2)}{(a-1)(a-2)} + k^2 \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)} \quad (11.7.40)$$

Simplifying and using part (a) gives part (b).

In the simulation of the beta-negative binomial experiment, vary the parameters and note the location and size of the mean \pm standard deviation bar. For various values of the parameters, run the simulation 1000 times and compare the empirical moments to the true moments.

This page titled [11.7: The Beta-Bernoulli Process](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.