Continuous Random Variables Lecture 5

The Beta distribution

The Beta distribution is a continuous distribution on the interval $(0,1)$. It is a generalization of the uniform random variable over $(0,1)$.

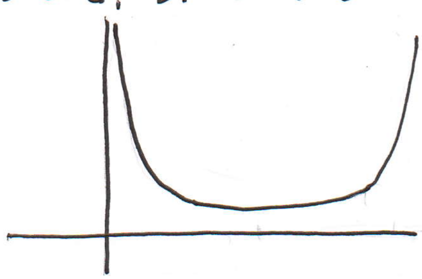
Def: A random variable X is said to have the Beta distribution with parameters a and b, where $a>0$ and $b>0$ if the pdf is

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1} \qquad 0<x<1$$

where the constant $\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$ is chosen to make the pdf integrate to 1. We write $X \sim Beta(a,b)$.

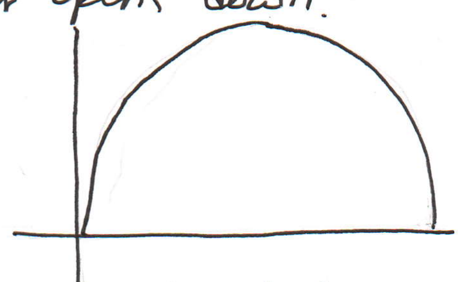
Observe that Beta$(1,1)$ is just the uniform distribution over $(0,1)$. In general

- If $a<1$ and $b<1$, the pdf is U-shaped and opens upward. If $a>1$ and $b>1$ the pdf opens down.



$a<1 \quad b<1$

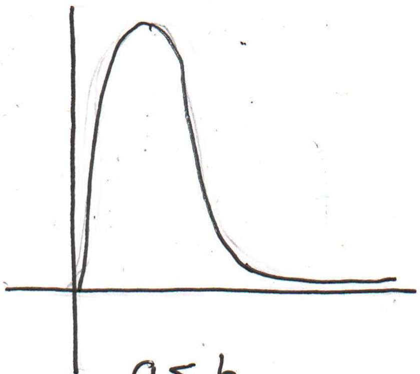
$(a=0.5,\ b=0.5)$



$a>1 \quad b>1$

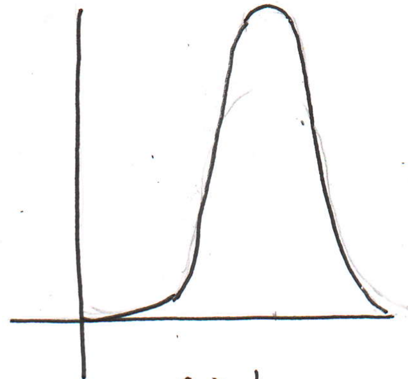
$(a=1.5,\ b=1.5)$

- If $a = b$ the pdf is symmetric about $\frac{1}{2}$. If $a > b$ the pdf favors values larger than $\frac{1}{2}$, if $a < b$, the pdf favors values smaller than $\frac{1}{2}$.



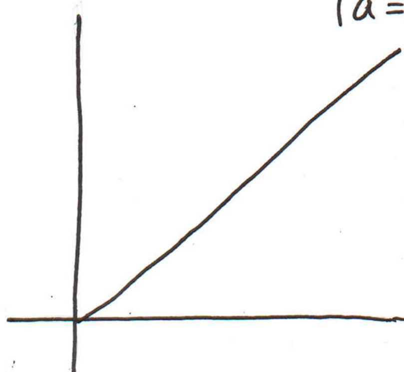
$a < b$

$(a = 2, b = 8)$



$a > b$

$(a = 2, b =$ [illegible]$)$



$a > b$

$(a = 2, b = 1)$

By definition, the constant $\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$.

In the special case where a and b are positive integers, Thomas Bayes figured out how to do the integral using a combinatorial argument rather than calculus.

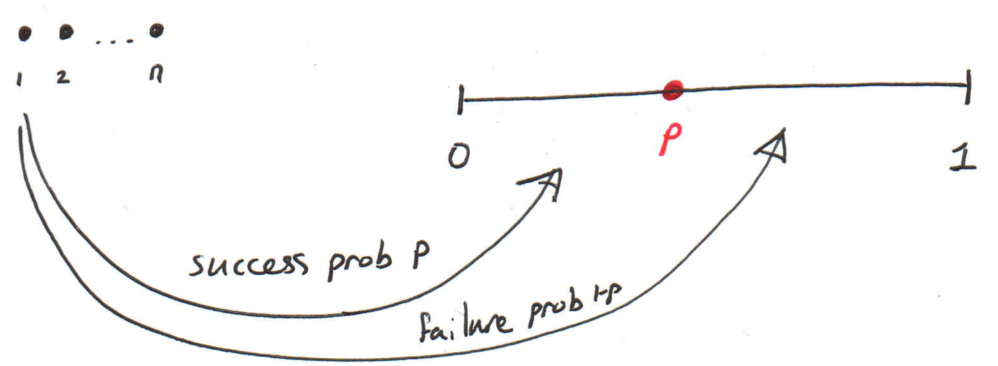
Proposition: $$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k}\, dx = \frac{1}{n+1}$$

for any integer $0 \leq k \leq n$.

Proof: Randomly pick $n+1$ points on the unit interval $[0,1]$. Color n of these points black and one point red. (You may imagine these points as atoms that are colored prior to being placed).

Define $X = 0, 1, \dots, n$ to be the number of atoms before the red one. If we first place the red particle at position $0 \leq p \leq 1$ and then assign randomly the position for all others we see that X becomes a binomial random variable with parameters (n, p). That is

$$P(X = k \,/\, \text{Red at } p) = P(X = k \mid R = p)$$
$$= \binom{n}{k} p^k (1-p)^{n-k}.$$



Since R is a uniform random variable with density function $f(p) = 1$, we can apply Bayes formula to compute $P(X=k)$ as follows:

Let $R_m = \frac{j}{m}$ $\quad 1 \leq j \leq m$ with $P(R_n = \frac{j}{m}) = f(\frac{j}{m})\frac{1}{m}$.

Then R_n is a discrete r.v. that is a good approximation for R when n is large.

Hence $P(X=k) = \sum_{j=1}^{m} P(X=k \mid R=\frac{j}{m}) P(R=\frac{j}{m})$

$$= \sum_{j=1}^{m} \binom{n}{k} \left(\frac{j}{m}\right)^k \left(1-\frac{j}{m}\right)^{n-k} f\left(\frac{j}{m}\right)\frac{1}{m} \longrightarrow \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} f(p)\,dp$$

$$= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\,dp$$

On the other hand, since the $n+1$ particles are placed randomly on the interval, it follows that all orderings of the particles are equally likely. That is, if the particles are labled by numbers $1, 2, \ldots, n, n+1$ where the $n+1$ particle is red, all $(n+1)!$ orderings are equally likely.

Thus $P(X=k) = \frac{n!}{(n+1)!} = \frac{1}{n+1}$

If particular, the above argument shows that if a and b are positive integers then

$$\int_0^1 \binom{a+b-2}{a-1} x^{a-1}(1-x)^{b-1}\,dx = \frac{1}{a+b-1}$$

Hence

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{(a-1)!\,(b-1)!}{(a+b-1)!}$$

The Beta distribution is a very useful tool in predicting the likelihood of an event from data gathered through experiments.

Ex. We have a coin that lands Heads with unknown probability p. Our goal is to infer the value of p after observing the outcomes of n tosses of the coin. The larger that n is, the more accurately we should be able to infer (or estimate) p.

If the coin is completely unknown to us, we might assume the default setting that p is equally likely to be any value between 0 and 1. That is, the default assumption that p is uniformly distributed seems natural and unbiased. More generally therefore, we may assume that the random variable $W = p$ has Beta distribution with parameters a, b. That is $W \sim \text{Beta}(a,b)$

(W stands for Wahrscheinlichkeit - probability in German)

Suppose we now perform a sequence of n independent tosses and find that k of them landed Heads. How should the pdf of W be updated?

Let X = # of successes (Heads). Given the pdf

$$f(p) = \frac{1}{\beta(a,b)} p^{a-1}(1-p)^{b-1} \text{ of } W,$$

we need to compute the marginal density function

$$g(p) = f|_{X=k}(p).$$

Perhaps the simplest way to go about it is to note that g must have the property $P(p \leq W \leq p+dp \mid X=k)$

$= g(p)dp$

* <u>Recall:</u> Probability at point p is the area of a thin rectangular column of height $g(p)$ and width dp.

Thus, $g(p)dp = P\left(p \leqslant W \leqslant p+dp \mid X=k\right)$

$$= \frac{P(X=k \mid p \leqslant W \leqslant p+dp)\, P(p \leqslant W \leqslant p+dp)}{P(X=k)}$$

Conditioning on W, $P(X=k) = \sum_{p}^{1} P\left(X=k \mid W \in [p, p+dp]\right) \cdot$
$\cdot P\left(W \in [p, dp+p]\right) \approx \binom{n}{k} \sum_{p} p^k (1-p)^{n-k} f(p)\, dp$

Hence $P(X=k) = \int_0^1 \binom{n}{k} x^k (1-x)^{n-k} f(x)\, dx.$

Similarly, $P(X=k \mid p \leqslant W \leqslant p+dp) \approx P(X=k \mid W=p)$

$= \binom{n}{k} p^k (1-p)^{n-k}$

and $P(p \leqslant W \leqslant p+dp) = f(p)\, dp$

Thus $g(p)dp = \dfrac{\binom{n}{k} p^k (1-p)^{n-k} f(p)\, dp}{\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} f(x)\, dx}$.

In particular $g(p) = C\, p^k (1-p)^{n-k} f(p) =$

$$= \tilde{C}\, p^k (1-p)^{n-k} p^{a-1} (1-p)^{b-1} \quad \text{where}$$

$$\tilde{C} = \frac{\binom{n}{k} \cdot \frac{1}{\beta(a,b)}}{\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} f(x)\,dx}$$

But $p^k (1-p)^{n-k} p^{a-1} (1-p)^{b-1} = p^{a+k-1} (1-p)^{b+n-k-1}$

and $\int_0^1 g(p)\,dp = 1$.

Thus $\tilde{C} = \dfrac{1}{\int_0^1 p^{a+k-1}(1-p)^{b+n-k-1}\,dp} = \dfrac{1}{\beta(a+k,\, b+n-k)}$

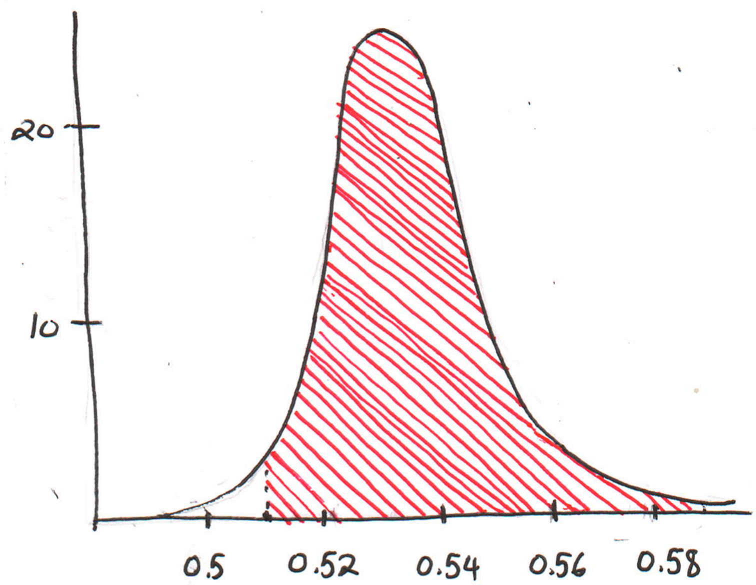
and $W|_{X=k} \sim \text{Beta}(a+k,\, b+n-k)$.

Ex. A coin is tossed 1000 times, 550 tosses come up Heads. Estimate the probability that the coin lands Heads.

Solution: A priori knowing nothing about the coin we estimate that $P(\text{Heads}) = p$ is uniformly distributed over $[0,1]$. Hence $p \sim \text{Beta}(1,1)$. Letting $X = \#$ of Heads we get $p|_{X=550} \sim \text{Beta}(1+550,\, 1+1000-550)$

$= \text{Beta}(551, 451)$

Using integration software, we see that the likelihood of the true probability of the coin to be a value between 0.51 and 0.59 is over 0.989.

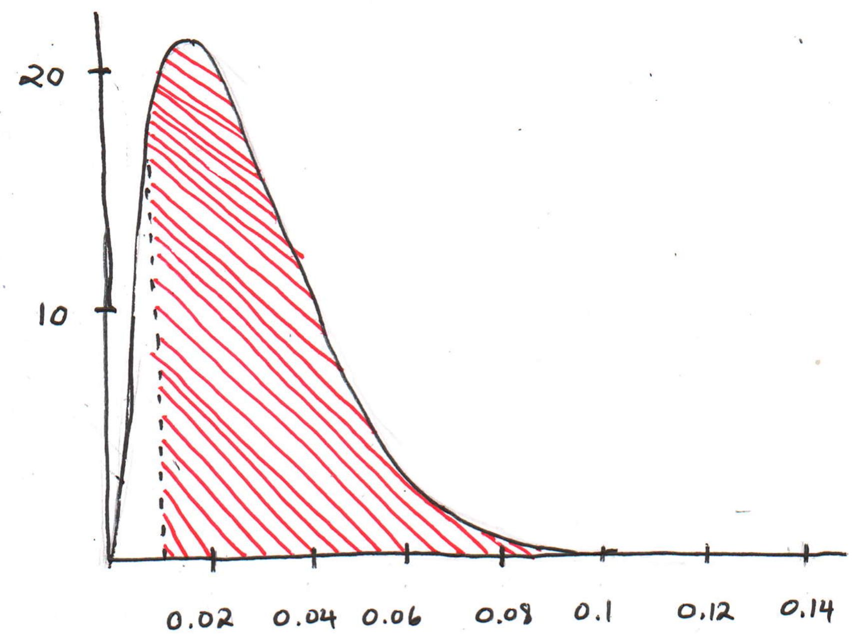


The likelihood this coin is fair is less than $\frac{1}{B(551,451)}\int_0^{0.51} p^{550}(1-p)^{450}\,dp$

≈ 0.00569

Ex. In a city of 12,000,000 inhabitants it is wished to estimate the number of people that contracted a certain disease. 100 subjects are randomly selected and carefully tested for the illness. Only 2 are found sick. What should be concluded about the prevalence of the disease in the general population?

<u>Solution:</u> Assuming the default probability is uniformly distributed (i.e. $p \sim Beta(1,1)$). The results of the study upgrade p to $Beta(1+2, 1+98) = Beta(3, 99)$



Using integration software, we see that the probability of the actual frequency being between 1% and 10% is 0.917

<u>Ex.</u> A new medicine against migrane is going through trials. n participants are admitted to the study.

(a) What is the probability that X, the number of participants whom the medicine helps, is equal to k?

(b) Given that all the participants in the study are positively affected, what is the probability that the true potentcy p of the drug is greater or equal to $\frac{1}{2}$?

Solution:

(a) By default, the probability p that drug is effective is $p \sim \text{Beta}(1,1)$.

Thus $P(X=k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \, dp = \frac{1}{n+1}$

(b) Given that drug worked on all n participants, we update $p \sim \text{Beta}(n+1, 1)$.

Now $\int_0^1 p^n \, dp = \frac{1}{n+1}$ so $\beta(n,1) = \frac{1}{n+1}$

Hence $P(p \geq \frac{1}{2}) = (n+1) \int_{\frac{1}{2}}^1 p^n \, dp = p^{n+1} \Big|_{\frac{1}{2}}^1 = 1 - \left(\frac{1}{2}\right)^{n+1}$