

Mathematics Stack Exchange is a question and answer site for people studying math at any level and professionals in related fields. Join them; it only takes a minute:

Here's how it works:

Sign up

Anybody can ask a question

Anybody can answer

The best answers are voted up and rise to the top



Sum of n i.i.d Beta-distributed variables

Ask Question



Assume that I have n variables that are each $X_i \sim \text{Beta}(\alpha, 1)$ distributed (with the same α , i.i.d.).

6

Is there anything known about the distribution of the sum $Y = \sum_i X_i$? I've tried to understand the works of Pham-Gia, but couldn't entirely grasp them. He seems to use some modified generalized Beta B^* .



★

An approximation is also fine with me, as long as I can compute some quantiles. I'm mostly interested in quantiles.

3

probability

probability-distributions

edited Nov 25 '11 at 16:43

asked Nov 25 '11 at 15:06



Anony-Mousse

387 1 4 16

Do you assume independent identically distributed variables X_i ? – Sasha Nov 25 '11 at 15:42

1 Yes, in the title I wrote "i.i.d." explicitly. – Anony-Mousse Nov 25 '11 at 16:42

2 Answers



Your Y will have mean $\frac{n\alpha}{\alpha+1}$ and variance $\frac{n\alpha}{(\alpha+1)^2(\alpha+2)}$

2

If you want a reasonable approximation other than a normal distribution with that mean and variance, you could say that $\frac{1}{n}Y$ has mean $\frac{\alpha}{\alpha+1}$ and variance $\frac{\alpha}{n(\alpha+1)^2(\alpha+2)}$ and then find a Beta



Home

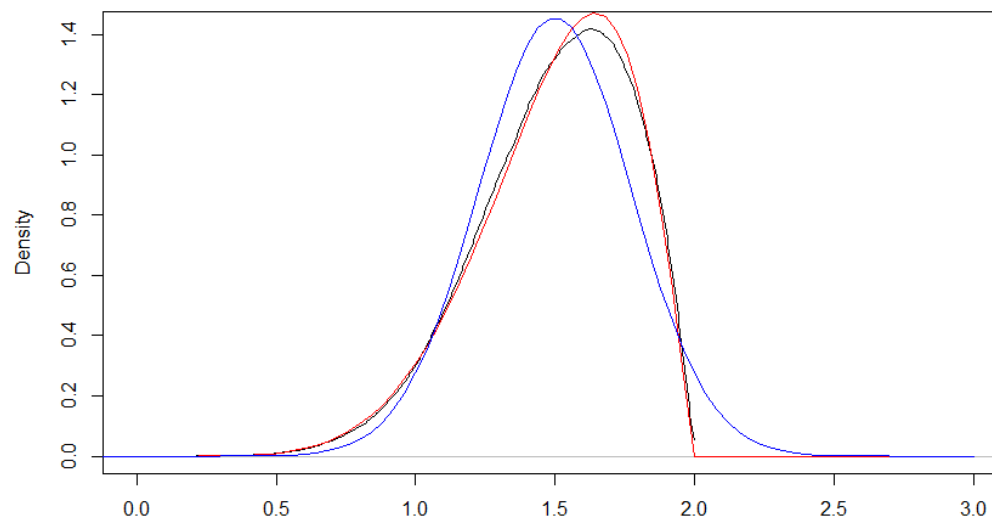
Questions

By using our site, you acknowledge that you have read and understand our Cookie Policy, Privacy Policy, and our Terms of Service.

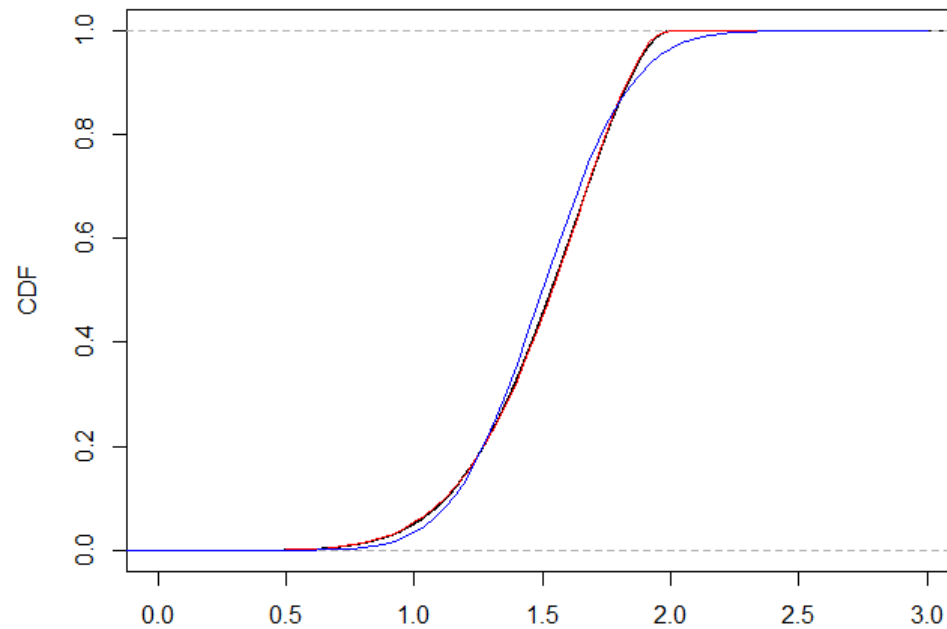


The density function for this approximation to $\frac{1}{n}Y$ would then be $\frac{\Gamma(\alpha' + \beta')}{\Gamma(\alpha')\Gamma(\beta')} x^{\alpha'-1} (1-x)^{\beta'-1}$ on $[0, 1]$ and the density function for the corresponding approximation to Y would be $\frac{\Gamma(\alpha' + \beta')}{n\Gamma(\alpha')\Gamma(\beta')} \left(\frac{x}{n}\right)^{\alpha'-1} \left(1 - \frac{x}{n}\right)^{\beta'-1}$ on $[0, n]$.

As an example, consider $\alpha = 3$ and $n = 2$. You get a mean of $\frac{3}{2}$ and variance of $\frac{3}{40}$, and for the approximation you get $\alpha' = \frac{27}{4}$, $\beta' = \frac{9}{4}$. In the following plot, the black curve gives the actual density, the red curve the stretched-Beta approximation and the blue curve the normal approximation. The red curve is not too far away in this case.



Similarly the cumulative distributions look like the next chart and you can see that the quantiles of the red stretched-Beta approximation are much closer than the quantiles of the blue normal approximation here.



answered Feb 2 at 1:21



Henry

101k 4 82 170

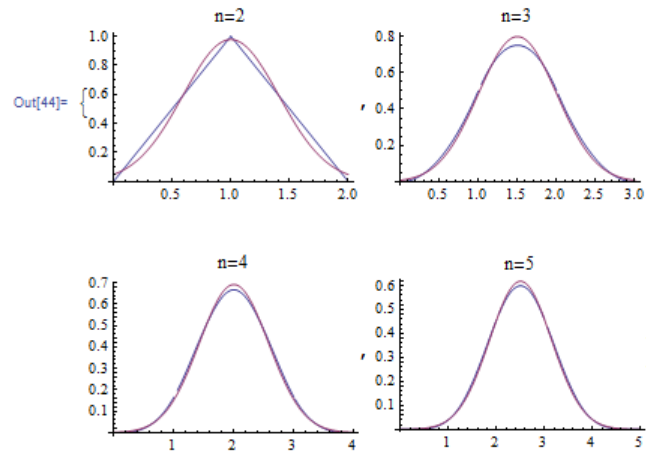
▲ I will assume X_i independent in this post. To get a feel for the problem, recall the special case of uniform distribution, corresponding to $\alpha = 1$, i.e. $\text{Beta}(1, 1) \stackrel{d}{=} U(0, 1)$.

3

▼ The sum of n iid uniform distribution was studied by J.O. Irwin and P. Hall, and the result is known as [Irwin-Hall distribution](#), aka uniform sum distribution.

Already for $n = 3$ the distribution density of the sum of three standard uniform variables approximates normal quite well:

```
In[44]= Table[Block[{di = UniformSumDistribution[n]},
  Plot[{PDF[di, x], PDF[NormalDistribution[Mean[di], StandardDeviation[di]], x]},
    {x, 0, n}, PlotLabel -> Row[{"n", "=", n}], {n, 2, 5}]
```



The same approximation will work well for larger values of n in your case as well. To write it out we need to compute mean and variance of the sum:

$$\mu_n = \mathbb{E}\left(\sum_{k=1}^n X_i\right) = n \frac{\alpha}{\alpha + 1} \quad \sigma_n^2 = \mathbb{V}\text{ar}\left(\sum_{k=1}^n X_i\right) = \sum_{k=1}^n \mathbb{V}\text{ar}(X_i) = \frac{n\alpha}{\alpha + 2} \frac{1}{(\alpha + 1)^2}$$

Thus the quantile function approximation is:

$$Q_n(q) \approx n \frac{\alpha}{\alpha + 1} + \frac{1}{\alpha + 1} \sqrt{\frac{n\alpha}{\alpha + 2}} Q_{\mathcal{N}(0,1)}(q)$$

For $n = 2$ CDF can be worked out exactly, and can be inverted using numerical algorithms:

```
In[47]= CDF[TransformedDistribution[x + y,
  {x ~ BetaDistribution[α, 1], y ~ BetaDistribution[α, 1]}], u]

Out[47]= {
  1, u ≥ 2
  (-1 + u)^α + u^2 α Beta[1/u, α, 1 + α] - u^2 α Beta[-1/u, α, 1 + α], 1 < u < 2
  (4^-α sqrt(π) u^2 α Gamma[α]) / Gamma[1/2 + α], 0 < u ≤ 1
  0, True
```

Added: The normal approximation can be truncated to $(0, n)$ interval to improve accuracy:

$$Q_{Y_n}(q) = \mu_n + \sigma_n Q_{N(0,1)}((1 - q)\Phi(-\mu_n/\sigma_n) + q\Phi((n - \mu_n)/\sigma_n))$$

edited Nov 25 '11 at 17:42

answered Nov 25 '11 at 16:24

"The same approximation will work well for larger values of n in your case as well" But probably not so well, is α is very different from 1. I would go this route, though, is n is large. – [leonbloy](#) Nov 25 '11 at 16:42

I need α to be different from 1, and n to be larger than 2. I'm particularly interested in $\alpha = .5$ (as this occurs in my situation for Euclidean distance, a prime example of an L_p norm - $1/p$ is actually my equation for α), and I'm interested in the development of this distribution when I increase n . Of course I do expect it to approximate a Gaussian distribution for large values of n ; at the same time I know it will never be *exactly* Gaussian, as negative values are not possible here. – [Anony-Mousse](#) Nov 25 '11 at 16:44 