EXERCISE: USING MC DRAWS FROM THE POSTERIOR

(1) Prove that if $\{p_1, p_2, p_3\}$ is distributed as Dirichlet with parameter vector $\{\alpha_1, \alpha_2, \alpha_3\}$, then the marginal distribution of $\{p_1/(p_1 + p_2), p_2/(p_1 + p_2)\}$ is Dirichlet(α_1, α_2). (Note that a two-dimensional Dirichlet is a Beta distribution)

This result can be generalized: If the vector $\vec{p}$ is distributed as Dirichlet$(\vec{\alpha})$, then any subvector of $\vec{p}$, normalized to sum to 1, has a Dirichlet marginal distribution, with parameter vector the corresponding subvector of $\vec{\alpha}$. This is also the conditional distribution of any subvector, normalized to sum to one, given the values of the other p_i's. The marginal distribution of any subvector $(p_1, \ldots, p_m)$ of $(p_1, \ldots, p_k)$, without rnormalization, is

$$\text{Dirichlet}(\alpha_1, \ldots, \alpha_m, 1 - \sum_1^m \alpha_j)$$

(2) In the AK dataset it appears that there is a downward drift, across year of birth cohorts, in the numbers of people who have zero years of schooling (`educ == 0`). What is the posterior probability that the population proportion of people with `educ==0` declines with every 2-year cohort (`yob` 30-30, 32-33, etc., so there are 5 cohorts)?

To compute this, the easiest way probably (unless you love doing many multiple integrals) is to draw a sample of, say, 1000 draws from the posterior distribution of the relevant population proportions, and count what fraction of them satisfy the condition.

The most straightforward approach is to use the fact that the population probabilities of the 5 cohort/zero education bins, normalized to sum to one across cohorts, are jointly 5-dimensional Dirichlet. Then it's just a matter of drawing a large number of 5-dimensional vectors and seeing how many of them satisfy `all(diff(p) < 0)`.

Note that this just checks whether the fraction of the total population that is accounted for by people with no education declines with 2-year cohort. We might instead want to know whether the fraction of the *cohort* that has no education is declining. The size of the cohorts does tend to increase over time, so this could affect results. In principle, the right way to handle this would be to draw both the probability of `educ==0` and the probability of `educ !=0` for each cohort and then work with the proportion of the cohort that has no education. However, because the number of people with some education is so much larger, posterior uncertainty of the ratio is totally dominated by the uncertainty about the proportion with no education.

Next most precise would be to recognize that since uncertainty about cohort size is relatively negligible (tens of thousands of draws instead of dozens), we could treat cohort sizes as constants and just scale the `educ == 0` probabilities by cohort size. However, this turns out to make no difference to the result.

Date: September 15, 2020.

R has no built-in function to generate random draws from a Dirichlet. A vector of independent Gamma variates, with shape parameters given by the vector α, when normalized to sum to one, is distributed as Dirichlet(α). Below is R code that uses this to make Dirichlet draws:

```
rdirichlet <- function(n, alpha) {
m <- length(alpha)
outmat <- matrix(0, n, m)
for (ic in 1:m) outmat[ , ic] <- rgamma(n, alpha[ic])
apply(outmat, 1, function(x) x / sum(x))
}
```