# Part 1

Throughout the 90's the General Social Survey gathered data about 155 women of age 40 at the time of the survey. These women were in their 20's during the 70's, which is considered a period of historically low fertility in the US. Assume that $Y_1, \ldots, Y_n \overset{iid}{\sim} \theta \sim \mathrm{Poisson}(\theta)$ denote the number of children that each of the n women have. Recall that for n observations assumed to come from a $\mathrm{Poisson}(\theta)$ generating model, the likelihood is $p(y_{1:n}|\theta) = \theta^{\sum y_i} e^{-n\theta} \prod_{i=1}^{n}(1/y_i!)\mathbf{1}_{\{y_i \in \mathbb{N}_0\}}$, which when assuming a $\mathrm{Gamma}(a,b)$ prior on $\boldsymbol{\theta}$ yields the posterior $p(\theta|y_{1:n}) \propto \theta^{\sum y_i + a - 1} e^{-\theta(n+b)} \mathbf{1}_{\{\theta>0\}} \implies \boldsymbol{\theta}|y_{1:n} \sim \mathrm{Gamma}\left(\sum y_i + a, n + b\right)$. Assume that $a = 2, b = 1$ for the prior of $\boldsymbol{\theta}$.

1. Using the information above, build a Monte Carlo algorithm to approximate the posterior predictive probability mass function (pmf). That is, sample indirectly $\tilde{Y}$ from the posterior predictive density as we did in page 7 of the Monte Carlo class notes (i.e., sampling first $\theta|y_{1:n}$ and then $\tilde{Y}|\theta$). Then use the sampled values to approximate $p(\tilde{y} = 0|y_{1:n}), p(\tilde{y} = 1|y_{1:n}), p(\tilde{y} = 2|y_{1:n}), \ldots$). Notice that a new observation $\tilde{Y}$ will have the same support as the data (i.e., $\mathbb{N}_0$); however, for large values of $\tilde{y}$ (say, $\tilde{y} = 20$, remember that the variable represents the number of children) the probabilities will be so close to zero that you will need to sample a very large number of times to actually draw samples of these values. Because of this, for large values of $\tilde{y}$ (for the sake of this exercise) can be assumed to be zero, so only approximate the values of the posterior predictive density for $\tilde{y} \in \{0, 1, 2, 3, \ldots, 10\}$. If in your samples you don't draw a particular number, the estimate for that probability is 0.

```r
#empirical distribution
ct.child <- rep(0,11)
ct.child[1:length(unique(w40$CHILDS))] <- as.vector(table(w40$CHILDS))
pr.child <- ct.child/sum(ct.child)
names(pr.child) <- names(ct.child) <- 0:10

#posterior predictive NegBinom(sum(y) + a, (b+n)/(b+n+1))
a <- 2; b <- 1
n <- nrow(w40)
sum.y <- sum(w40$CHILDS)
p <- (b+n)/(b+n+1)
post.pred <- dnbinom(0:10,size=(a+sum.y),prob=p) # mu=((a+sum.y)/(b+n)))#

#approximate predictive Gamma(sum(y)+a,b+n)-Poisson(theta.k) mixture
S <- 20000
theta.k <- ytilde.ks <- NULL
for(k in 1:S){
  theta.k <- rgamma(1,a+sum.y,b+n)
  ytilde.ks <- rpois(n,theta.k)
}
appr.child <- rep(0,11)
appr.child[1:length(unique(ytilde.ks))] <- as.vector(table(ytilde.ks))
appr.pred <- appr.child/sum(appr.child)
names(appr.pred) <- names(appr.child) <- 0:10
```
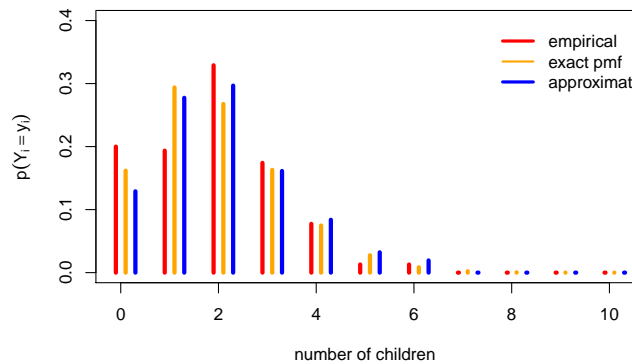
Compared with empirical values, the approximate posterior predictions have closed results.

But each run gives a different result. Sometime underestimate the probabilities, sometimes overestimate. Meanwhile, the exact posterior prediction gives a non-dithering result.

2. Show that the actual posterior predictive distribution $p(\tilde{y}|y_{1:n})$ for a new observation under this particular setup is $\text{NegBinom}\left(\sum y_i + a, (b+n)/(b+n+1)\right)$. The negative binomial pmf can be represented in two different ways, so for this problem assume that a random variable $X|r,\theta \sim \text{NegBinom}(r,\theta)$ if its pmf is given by $p(x|r,\theta) = \binom{x+r-1}{x}(1-\theta)^r\theta^x$, $r > 0$, $\theta \in (0,1)$, for $x \in \{0,1,2,\dots\}$.

$$p(\tilde{y}|y_{1:n}) = \int_{-\infty}^{\infty} p(\tilde{y}|\theta)p(\theta|y_{1:n})d\theta$$

$$= \int_{-\infty}^{\infty} Pois(\tilde{y}|\theta)Gamma(\theta|\sum y_i + a, n+b)d\theta$$

$$= \int_{-\infty}^{\infty} \mathbf{1}_{\{\tilde{y}\in\mathbb{N}_0\}}(1/\tilde{y}!)\theta^{\tilde{y}}e^{-\theta}\frac{(n+b)^{\sum y_i+a}}{\Gamma(\sum y_i + a)}\theta^{\sum y_i+a-1}e^{-\theta(n+b)}\mathbf{1}_{\{\theta>0\}}d\theta$$

$$= \mathbf{1}_{\{\tilde{y}\in\mathbb{N}_0\}}\frac{(n+b)^{\sum y_i+a}}{\tilde{y}!\Gamma(\sum y_i + a)}\int_0^{\infty}\theta^{(\tilde{y}+\sum y_i+a-1)}e^{-\theta(n+b+1)}d\theta$$

$$= \mathbf{1}_{\{\tilde{y}\in\mathbb{N}_0\}}\frac{(\tilde{y}+\sum y_i+a-1)!}{\tilde{y}!(\sum y_i+a-1)!}\frac{(n+b)^{\sum y_i+a}}{(n+b+1)^{(\tilde{y}+\sum y_i+a)}}\int_0^{\infty}\frac{(n+b+1)^{(\tilde{y}+\sum y_i+a)}}{\Gamma(\tilde{y}+\sum y_i+a)}\theta^{(\tilde{y}+\sum y_i+a-1)}e^{-\theta(n+b+1)}d\theta$$

$$= \mathbf{1}_{\{\tilde{y}\in\mathbb{N}_0\}}\binom{\tilde{y}+\sum y_i+a-1}{\tilde{y}}(1-\frac{1}{n+b+1})^{\sum y_i+a}(\frac{1}{n+b+1})^{\tilde{y}}$$

The posterior predictive density for $\tilde{Y}$ is $\text{NegBinom}(\tilde{y}|\sum y_i + a, \frac{1}{b+n+1})$.

3. Compare graphically the quality of the approximation to the exact pmf for $S = 2000, 5000, 10000$ (the number of Monte Carlo samples) and comment on your results.

```
S <- 2000
theta.k <- ytilde.ks <- NULL
for(k in 1:S){theta.k <- rgamma(1,a+sum.y,b+n); ytilde.ks <- rpois(n,theta.k)}
appr.child <- rep(0,11)
appr.child[1:length(unique(ytilde.ks))] <- as.vector(table(ytilde.ks))
appr.pred.2k <- appr.child/sum(appr.child)
names(appr.pred.2k) <- names(appr.child) <- 0:10
```

```
S <- 5000
theta.k <- ytilde.ks <- NULL
for(k in 1:S){theta.k <- rgamma(1,a+sum.y,b+n); ytilde.ks <- rpois(n,theta.k)}
appr.child <- rep(0,11)
appr.child[1:length(unique(ytilde.ks))] <- as.vector(table(ytilde.ks))
appr.pred.5k <- appr.child/sum(appr.child)
names(appr.pred.5k) <- names(appr.child) <- 0:10

S <- 10000
theta.k <- ytilde.ks <- NULL
for(k in 1:S){theta.k <- rgamma(1,a+sum.y,b+n); ytilde.ks <- rpois(n,theta.k)}
appr.child <- rep(0,11)
appr.child[1:length(unique(ytilde.ks))] <- as.vector(table(ytilde.ks))
appr.pred.10k <- appr.child/sum(appr.child)
names(appr.pred.10k) <- names(appr.child) <- 0:10
```
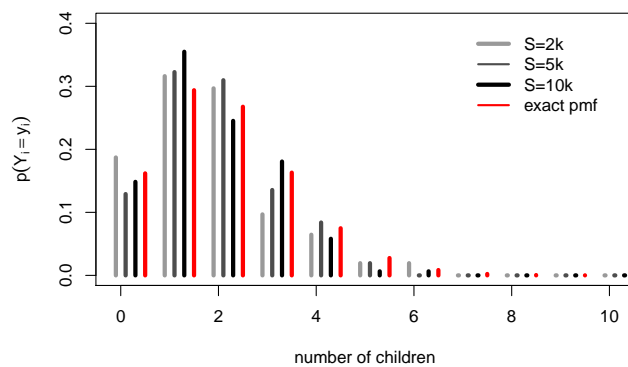

Comparing the quality of the approximation to the exact pmf for different number of Monte Carlo samples, the larger number shows a closer approximation to the exact pmf. But this process didn't show a stable prefermance and may need a longer sequence to confirm the trend.

Part 2

4. Suppose $X_1, X_2, \ldots, X_n \overset{iid}{\sim} p(x|\theta)$ given θ. Imagine that the posterior predictive for a new observation $\tilde{X}$, $p(\tilde{x}|x_{1:n})$, is too complicated to compute analytically; however, you can easily sample from the posterior $p(\theta|x_{1:n})$, and you can also easily compute the cdf $F(X \leq a|\theta)$. Construct a Monte Carlo estimator for $\Pr(\tilde{X} < a|x_{1:n})$ for some particular value of a.

Use the Monte Carlo method to approximate a posterior predictive density. The posterior predictive density for a new observation $\tilde{X}$ is defined as

$$p(\tilde{x}|x_{1:n}) = \int_{\theta \in \Theta} p(\tilde{x}|\theta)p(\theta|x_{1:n})d\theta = E[p(\tilde{x}|\vec{\theta})|x_{1:n}] \approx \frac{1}{S}\sum_{k=1}^{S} p(\tilde{x}|\vec{\theta}_k)$$

where $\vec{\theta}_1, \ldots, \vec{\theta}_S \overset{iid}{\sim} p(\theta|x_{1:n})$.
Sampling θ_k from $p(\theta|x_{1:n})$ then sampling $\tilde{x}_k$ from $p(\tilde{x}|\theta_k)$ can generate samples from $p(\tilde{x}|x_{1:n})$ for $k = 1, 2, \ldots, S$

This yields the sequence of S independent bivariate samples $[(\theta_1, \tilde{x}_1), \ldots, (\theta_S, \tilde{x}_S)]$ drawn from the joint posterior distribution of $(\vec{\theta}, \tilde{X})$ given $X_{1:n} = x_{1:n}$. Dropping the θ_k's, the values $[\tilde{x}_1, \ldots, \tilde{x}_S]$ constitute S

independent draws from the (marginal) posterior predictive $p(\tilde{x}|x_{1:n})$. This is equivalent to integrating out $\vec{\theta}$.

When X has cdf F_X, the random variable $F_X^{-1}(U)$, where $U \sim uniform(0,a)$, has distribution F_X. $F_X^{-1}(u) = x \leftrightarrow \int_{-\infty}^{x} f_x(y)dy$

$$\Pr(\tilde{X} < a|x_{1:n}) = \int_0^a p(\tilde{x}|\theta)dx = a \int_{-\infty}^{\infty} p(\tilde{y}|\theta)(\frac{1}{a}\mathbf{1}_{\{y \in (0,a)\}})dy$$

$$= aE(g(Y)) \approx \frac{a}{S}\sum_{k=1}^{S} g(Y_k)$$

where $g(Y) = p(x|\theta)$ and $Y \sim U(0,a)$ (i.e., has pdf $p(y) = \frac{1}{a}\mathbf{1}_{\{y \in (0,a)\}}$).