Week 1 HW 2019122064 DagunOh

Part 2. HW

1) Part 의 code를 수정해서 (Strong likelihood, Weak likelihood) $\times$ (Uninformative prior, Weak prior, Strong prior)의 6가지 경우 비교해보기

2) BDA 1.3 Exercise
Suppose that in each individual of a large population there is a pair of genes, each of which can be either x of X, that controls eye color: those with xx have blue eyes, while heterozygotes (those with Xx or xX) and those with XX have brown eyes. The proportion of blue-eyed individuals is p^2 and of heterozygotes is $2p(1-p)$, where $0<p<1$. Each parent transmits one of tis own genes to the child; if a parent is a heterozygote, the probability that it transmits one of its own genes to the child; if a parent is a heterozygotes, the probability that it transmits the gene of type X is $\frac{1}{2}$. `Assuming random mating, show that among brown-eyed children of brown-eyed parents, the expected proportion of heterozygotes is` $\frac{2p}{1+2p}$. Suppose Judy, a brown-eyed child of brown-eyed parents, marries a heterozygote, and they have n children, all brown-eyed. `Find the posterior probability that Judy is a heterozygote and the probability that her first grandchild hs blue eyes.`

3) 새로운 대학병원에서의 high risk 수술의 생존율에 관한 분석. 다른 병원에서의 경험을 통해 생존율은 0.9 정도로 예상되며 0.8 미만이거나 0.97 초과일 것 같지는 않다고 생각한다.

3-a) \textbf{Beta} distribution으로 위의 \textbf{belief}을 survival rate에 관한 \textbf{prior distribution}으로 나타내라. Parameter α, β는 어떻게 선정하면 좋을 것인가?
(Hint : 여러분의 믿음의 강도 따라 α, β의 답이 달라질 수 있다. 하나의 정답을 맞추는 것이 아니라 실생활의 문제를 해석하는 힘을 기르는 것이 취지라 하겠다.)

3-b) 이제 \textbf{data gathering}. 10명의 환자에 수술을 진행해 모두 생존하였다. survival rate에 관한 \textbf{Posterior Distribution} 구하기.

3-c) 다음 환자가 생존할 확률과 다음 20명의 환자 중 2명 이상 사망할 확률을 각각 예측하시오. (Hint : Posterior Predictive)

---

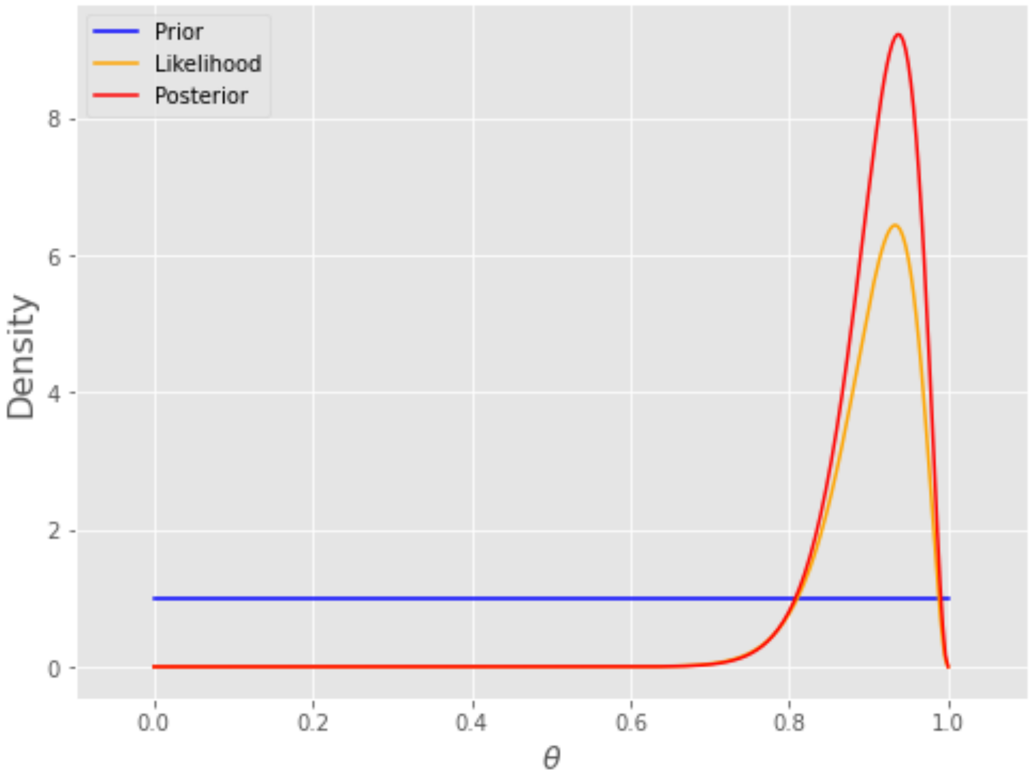
1. 6 cases comparison

```python
def likelihood(theta, n, y):
    return theta**y*(1-theta)**(n-y)
```

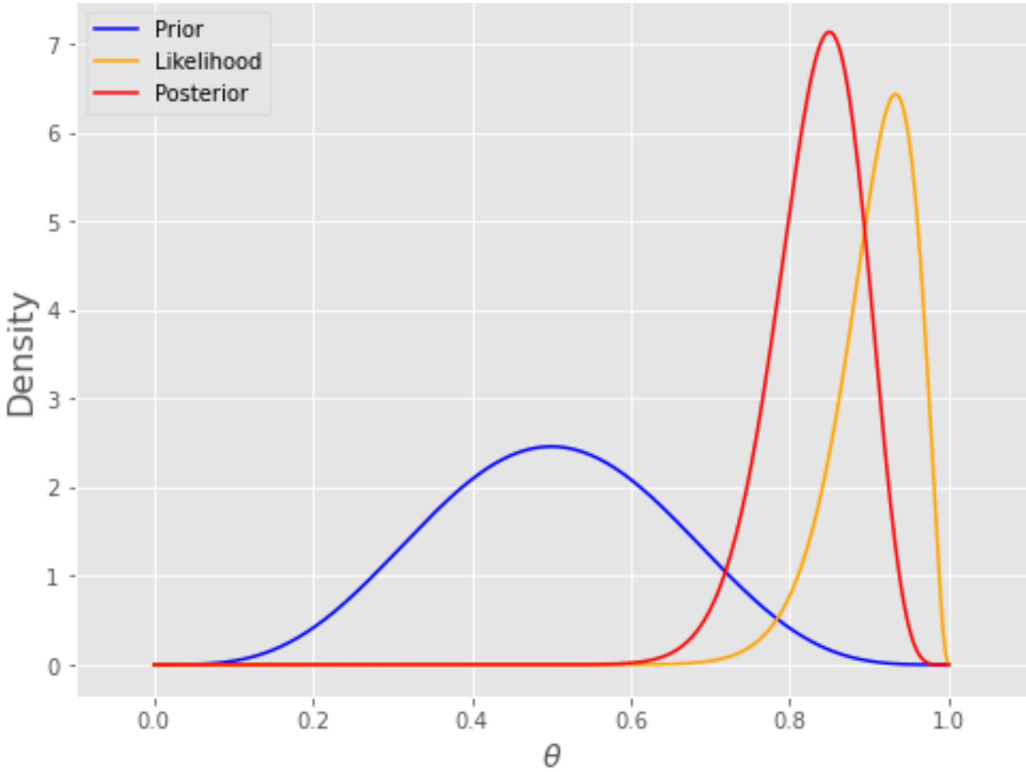
```python
def comparisonplot(a0,b0,n,y):
    # Prior : variable
    # uninformative : p(theta) = 1
    prior = st.beta(a=a0, b=b0)
```

```python
    # Posterior
    post = st.beta(a=a0+n, b=b0+(n-y))

    # plotting
    thetas = np.linspace(0, 1, 300)
    plt.figure(figsize=(8, 6))
    plt.style.use('ggplot')
    plt.plot(thetas, prior.pdf(thetas), label='Prior', c='blue')

    # 원래는 likelihood 앞에 막 이렇게 곱하면 안되지만 존재하는 것 알려주기 위해.
    plt.plot(thetas, (10**4)*likelihood(thetas, n, y), label='Likelihood',
c='orange')
    plt.plot(thetas, post.pdf(thetas), label='Posterior', c='red')
    plt.xlim([-0.10, 1.10])
    plt.xlabel(r'$\theta$', fontsize=14)
    plt.ylabel('Density', fontsize=16)
    plt.legend();
```

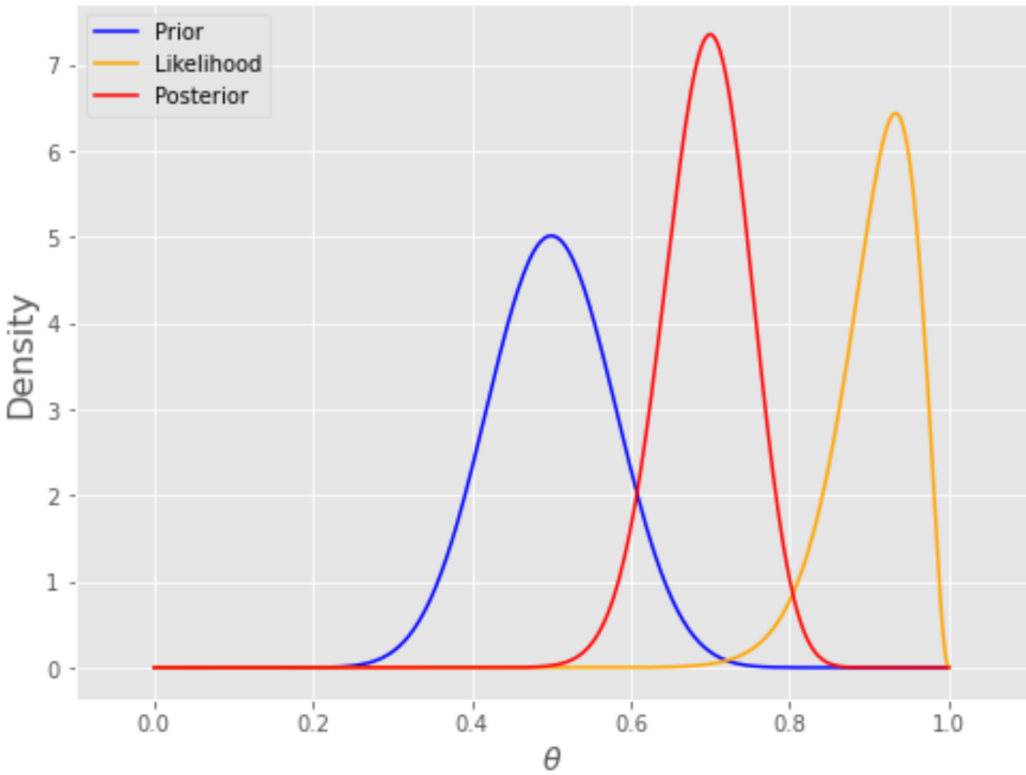
```python
#strong likelihood X Uninformative prior
comparisonplot(1,1,30,28)
```



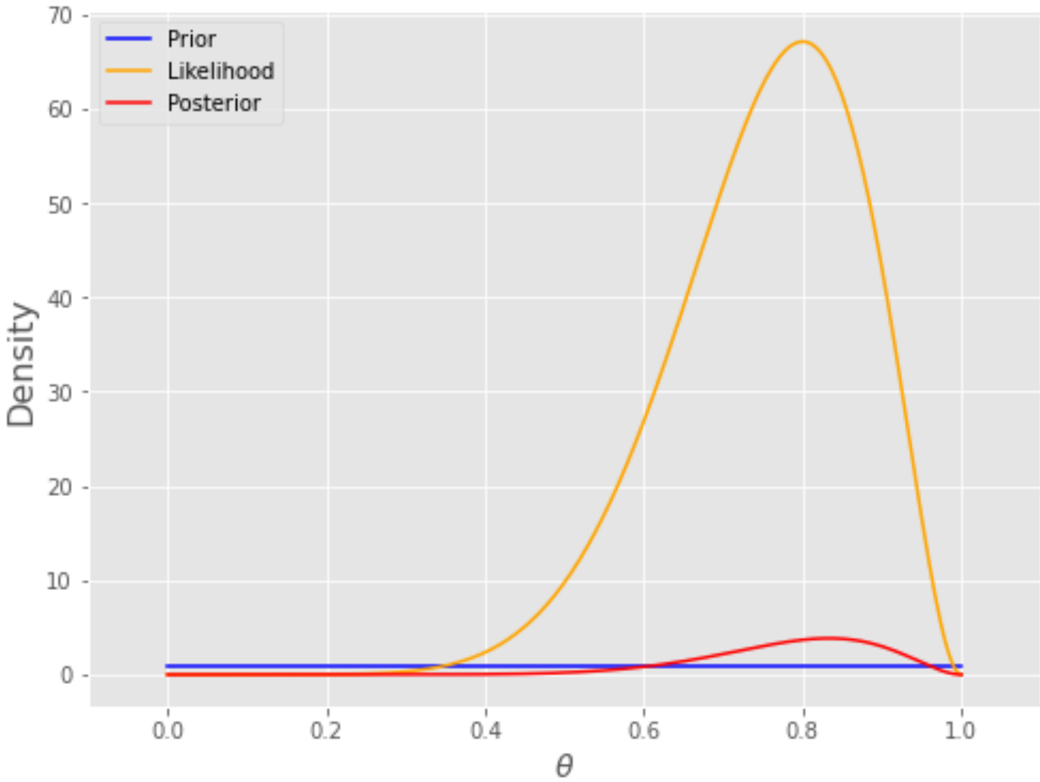

```python
#Strong likelihoood X weak prior
comparisonplot(5,5,30,28)
```



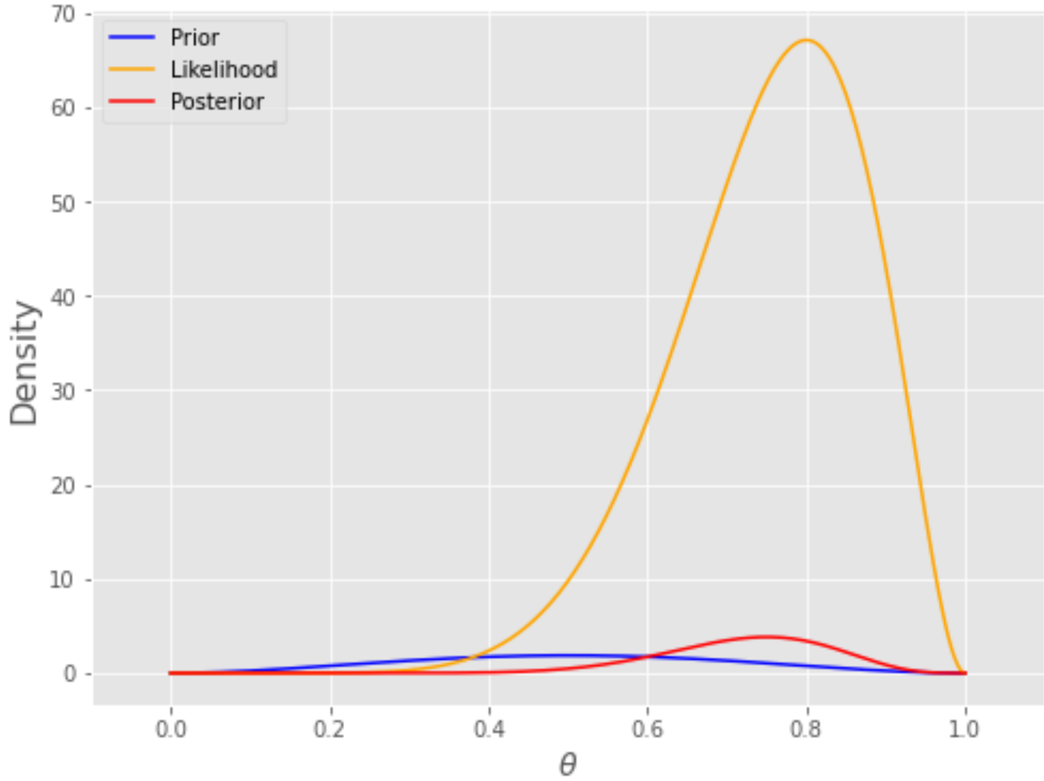

```
#Strong likelihood X strong prior
comparisonplot(20,20,30,28)
```



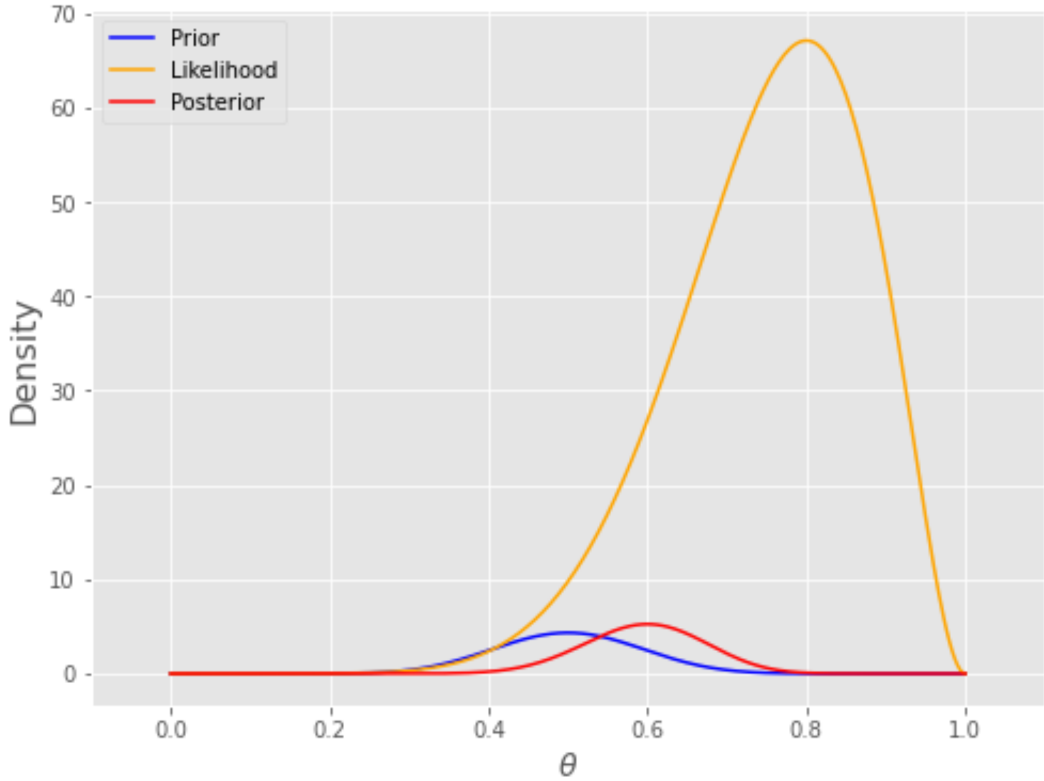
```
#Weak likelihood X uninformative prior
comparisonplot(1,1,10,8)
```




```
#Weak likelihood X weak prior
comparisonplot(3,3,10,8)
```




```
#Weak likelihood X strong prior
comparisonplot(15,15,10,8)
```



2.

xx : 열성 유전자 : blue eyes : p^2

Xx , xX , XX : 우성유전자 X로 인한 brown eyes

Xx 이 X 물려줄 확률 $\frac{1}{2}$.

잡종 : $2p(1-p)$

① Show brown-eyed children of brown-eyed parents is $\frac{2p}{2p+1}$

<proportion>

$p(x) = p \quad p(X) = (1-p) \rightarrow p(xx) = p^2, \quad p(Xx) = p(1-p), \quad \underset{\text{순수}}{p(XX)} = (1-p)^2$

$$p(\overset{A}{\text{heterozygotes}} \mid \overset{B}{\text{brown-eyed child \& brown eyed parents}})$$

$$= \frac{p(A, B)}{p(B)}$$

$$= \frac{p(\text{heterozygote} \overset{\text{''}D}{\text{brown eyed child}}, Xx \cdot xX) + p(D, XX \cdot XX) + p(D, XX \cdot Xx) \times 2}{p(\underset{\underset{C}{\text{''}}}{\text{brown eyed child}}, Xx \cdot Xx) + p(C, XX \cdot XX) + p(C, Xx \cdot XX) \times 2}$$

부·모 유전자 반대.

$$= \frac{\frac{2}{3} \times \frac{3}{4} \times 2p(1-p) \times 2p(1-p) + \overset{(\text{잡종 X})}{0} + 2 \times \frac{1}{2} \times 2p(1-p) \times (1-p)^2}{\frac{3}{4} \times 2p(1-p) \times 2p(1-p) + 1 \times (1-p)^2 (1-p)^2 + 2 \times (1-p)^2 2p(1-p)}$$

$$= \frac{\frac{1}{2} \times 4p^2 + 2p(1-p)}{3p^2 + 4p(1-p) + (1-p)^2} = \frac{\cancel{2p^2} + 2p - \cancel{2p^2}}{\cancel{3p^2} + \cancel{4p} - \cancel{4p^2} + \cancel{p^2} - \cancel{2p} + 1}$$

$$= \frac{2p}{2p+1}$$

$$\therefore \frac{2p}{2p+1}$$

↳ brown → brown $\left(\frac{2p}{2p+1}\right)$

② $P(\text{Judy\_hetero} \mid \text{Judy} \times \text{hetero}\ \&\ \text{brown children})$

$$= \frac{\overset{①}{P(\text{Judy\_hetero})}}{P(\text{Judy} \times \text{hetero}\ \&\ \text{brown children})}$$

① $P(\text{Judy\_hetero} \times \text{hetero} \sim)$ + ③ $P(\text{Judy\_homo} \times \text{hetero} \sim)$

XX Xx

$$= \frac{\cancel{\frac{2p}{2p+1}} \times \cancel{2p(1-p)} \times \left(\frac{3}{4}\right)^n}{\cancel{\frac{2p}{2p+1}} \times \cancel{2p(1-p)} \times \left(\frac{3}{4}\right)^n + \cancel{\frac{1}{2p+1}} \times \cancel{2p(1-p)} \times 1}$$

Judy hetero 여사건

↳ 무조건 brown eyes

$$= \frac{2p\left(\frac{3}{4}\right)^n}{2p\left(\frac{3}{4}\right)^n + 1} = p'$$

③ p(first grandchild blue eyed | Judy × hetero & all brown children)

$(1-p')$ Judy_homo Judy_hetero / * if parents hetero. $\frac{2p}{1+2p}$

XX Xx

grand child 가 blue eyes 나오려면 child 무조건 hetero

Judy 배우자

Xx × Xx

child ~~XX~~ Xx xX × 배우자 xx → $\frac{1}{2}$

~~XX xx~~ Xx → $\frac{1}{4}$

Xx xx

Xx Xx xx xx

XX ~~Xx~~

~~XX~~ Xx Xx ~~XX~~ × xx $\frac{1}{2}$

× Xx $\frac{1}{4}$

④ $p' \times \underset{\text{child}}{\frac{2}{3}} \times \left(\underset{\text{hetero 배우자}}{\frac{1}{4} \times 2p(1-p)} + \underset{\text{열성 homo 배우자}}{\frac{1}{2} \times p^2}\right) + (1-p') \times \frac{1}{2} \times \left(\frac{1}{4} \times 2p(1-p) + \frac{1}{2}p^2\right)$

3번)

경험통한 생존률 0.9

0.8 미만이거나 0.97 초과는 아닐것 같다고 생각.

a) $p(\theta) \sim Beta(\alpha, \beta)$ → 생존률 경험상 0.9 → 평균 $\frac{a}{a+b} = 0.9$ $a=45, b=5$

$$0.8 \le \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} \fallingdotseq 0.9 < 0.97$$

b) 10명 수술해서 모두 생존.

posterior distribution ~ Beta (a+y, n+b-y)

conjugacy 이용. binomial 한 상황

$$E(\theta|y) = \frac{a+y}{a+b+n} \rightarrow \frac{55}{50+10} \quad y=5.$$

$a+y = 55, \quad n+b-y = 5.$

posterior distribution ~ Beta(55, 5)

c)

i) 다음 환자가 생존할 확률.

$$P(X_2=1 | X_1=10) = \int_0^1 P(X_2=1 | X_1=10, \theta=\theta)\, p(\theta)\, d\theta$$

$$= \int_0^1 P(X_2=1 | \theta=\theta) \cdot f_{\theta|X_1=10}(\theta)\, d\theta = \frac{\beta(56,5)}{\beta(55,5)} = 0.9166..$$

2, 3, 4 ... 20 : 1 − P(1 or 0 death)

ii) 다음 20명의 환자 中 2명 이상 사망할 확률.

1 − P(0 or 1 deaths)

$= 1 - P(0\ death) - P(1\ death)$

$$= 1 - \int_0^1 \theta^{20} \cdot \frac{\theta^{54} \cdot (1-\theta)^4}{\beta(55,5)}\, d\theta - \int_0^1 {}_{20}C_1 \cdot \theta^{19}(1-\theta) \cdot \frac{\theta^{54}(1-\theta)^4}{\beta(55,5)}$$

$$= 1 - \int_0^1 \frac{\theta^{74}(1-\theta)^4}{\beta(55,5)} \, d\theta - \int_0^1 {}_{20}C_1 \frac{\theta^{73}(1-\theta)^5}{\beta(55,5)}$$

$$= \quad 1 - \frac{\beta(75,5)}{\beta(55,5)} \quad - 20 \times \frac{\beta(74,6)}{\beta(55,5)} \quad = \quad 0.47768\cdots$$