Синонимия гамма распределений

02.08.2019

Модели гамма распределения и их приложения

Одной из распространенных проблем в статистическом анализе биометрических данных является ограничение возможностей методов, основанных на проверке гипотез о равенстве средних. Особенно актуально это становится при исследовании данных типа "мышца" – выборочных наблюдений, в которых значимый вклад в дисперсию вносят факторы сжатия-растяжения.

Например, как данные типа "мышца" можно рассматривать результаты морфологических измерений высоты клеток секреторного эпителия альвеолы молочной железы (лактоцитов) у лактирующих мышей, так как форма этих клеток постоянно меняется в зависимости от стадии выработки секрета.

В двух экспериментах по исследованию морфологии лактоцитов при введении лабораторным животным различных по воздействию на лактацию препаратов – простагландина $F_{2\alpha}$ и тиреолиберина – были обнаружены одинаково значимые эффекты увеличения высоты эпителия по сравнению с контролем, т. е. проверка гипотез о равенстве средних по сути оказалась бесполезной. В такой ситуации продолжение исследования обычно сопровождается проверкой гипотез об одинаковости функций распределения и при значимом их отличии – поиском адекватной статистической модели.

Вид гистограмм распределения высоты секреторного эпителия при разных условиях эксперимента позволили предположить, что должно быть согласие с гамма распределением.

Необходимые функции для ОМП гамма распределения.

```r
library(bbmle)
```

```
## Loading required package: stats4
```

```r
library(hypergeo)

EstGamma<-function(x)
{
  minusLog<-function(shape,rate,x)
    return(-sum(dgamma(x,shape=shape,
                       rate=rate,
                       log=TRUE)))

  M<-mean(x)
  V<-var(x)
  alpha<-M/V
  lambda<-M*alpha


  file<-as.numeric(x)

    m <- mle2(
    minuslogl = minusLog,
    start = list(shape = lambda, rate = alpha),
    data = list(x = file),
    "L-BFGS-B",
    lower = list(shape = 1e-4, rate = 1e-6),
    upper=list(shape = 1e+5, rate = 1e+5))
```

```
  shape <- coef(m)[1]
  rate <- coef(m)[2]

list(c(shape=shape,rate=rate),m)
}
```

Степенное гамма распределение

Найдем распределение случайной величины $\eta = \xi^{\frac{1}{\kappa}}$, где κ положительное число, ξ гамма распределенная $\gamma(\alpha, \lambda)$ случайная величина с плотностью

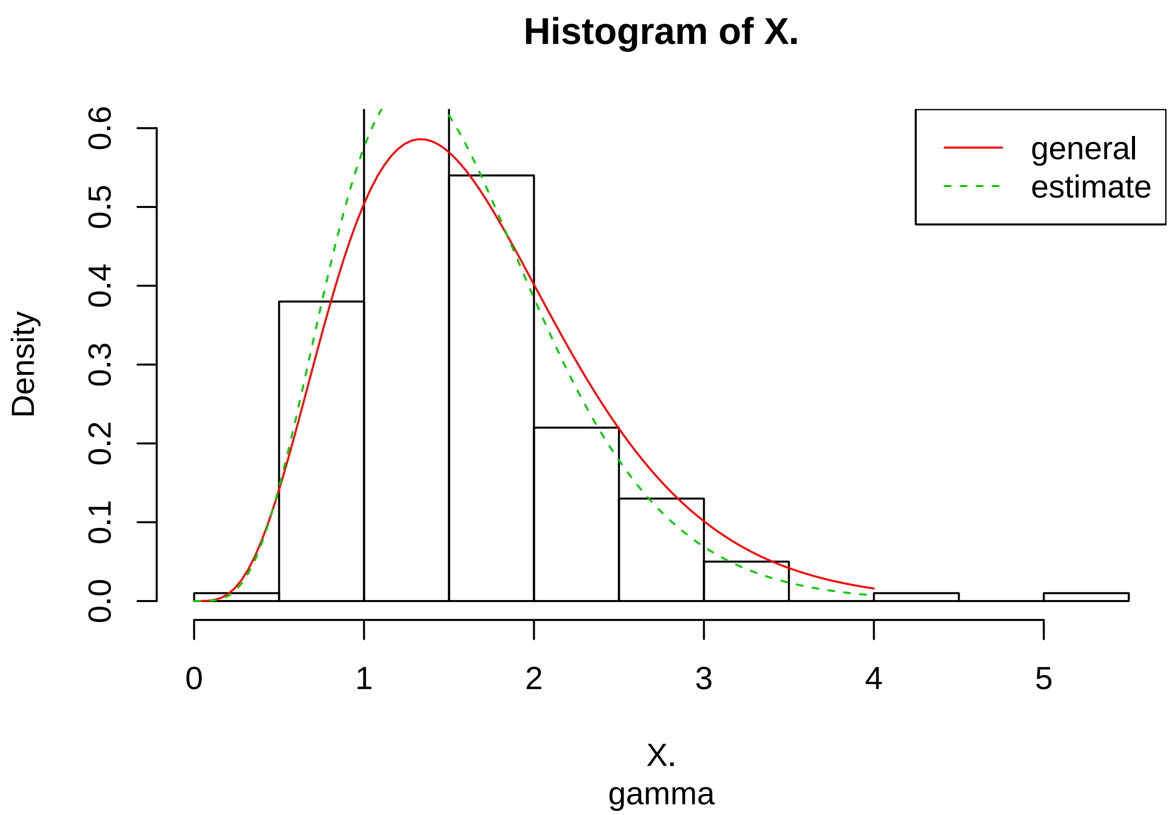
$$\gamma(x|\alpha, \lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}\,.$$

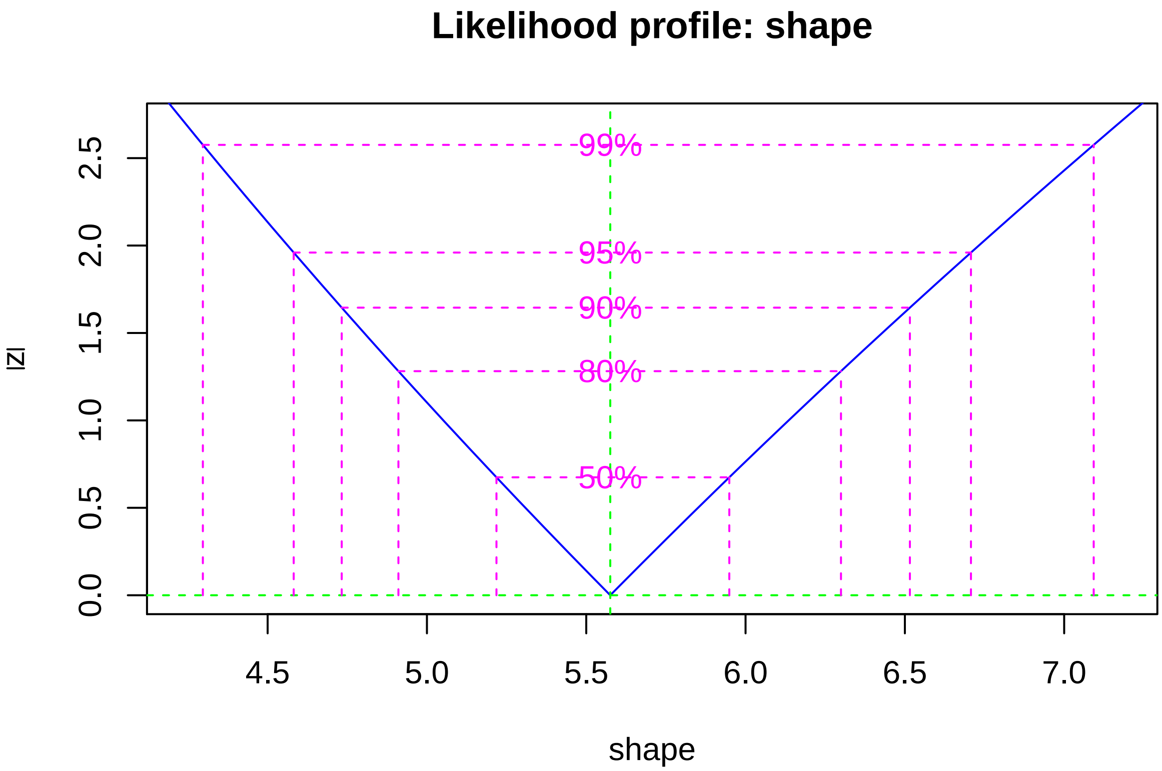
Вычислим сначала функцию распределения.

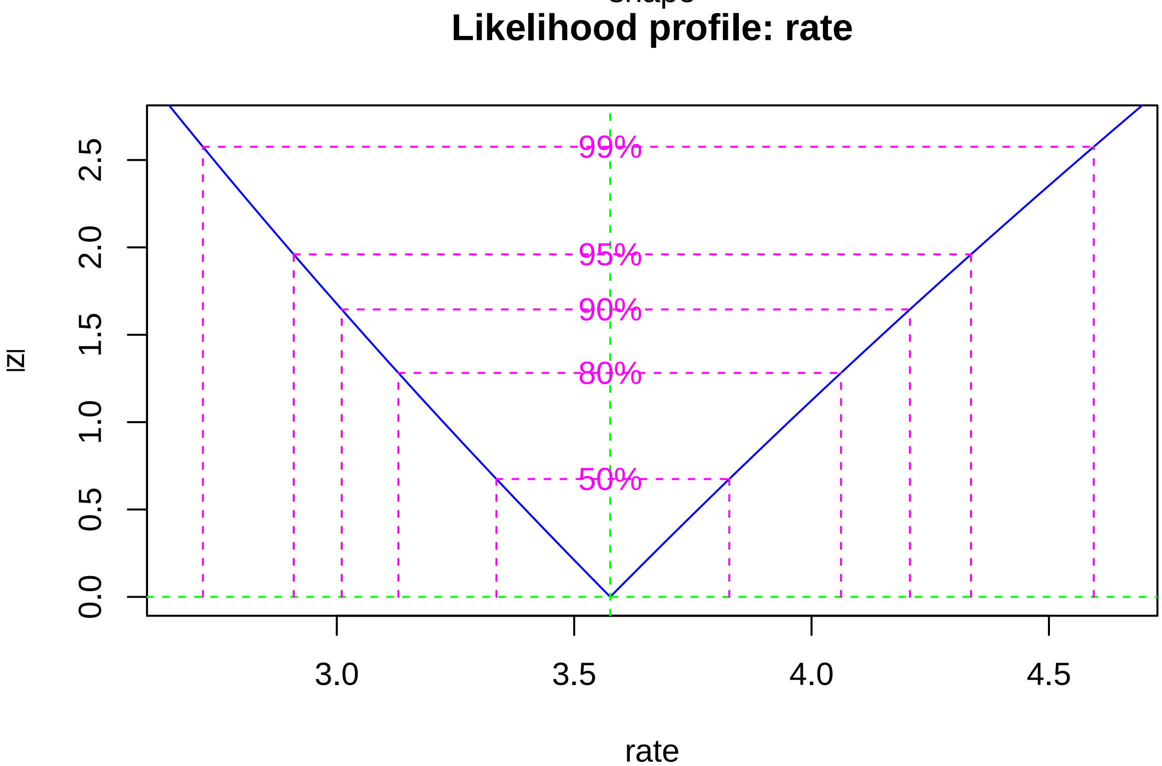
$$P\{\xi^{\frac{1}{\kappa}} < x\} = P\{\xi < x^\kappa\} = F_\xi(x^\kappa),$$

$$f_\eta(x) = \kappa x^{\kappa-1} f_\xi(x^\kappa) = \kappa x^{\kappa-1} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\kappa(\lambda-1)} e^{-\alpha x^\kappa} = \frac{\kappa\alpha^\lambda}{\Gamma(\lambda)} x^{\kappa\lambda-1} e^{-\alpha x^\kappa}\,.$$

Промоделируем $\gamma(3, 5)$ и сравним гистограмму с заданной плотностью распределения. Можно построить график плотности распределения с оценками параметров.

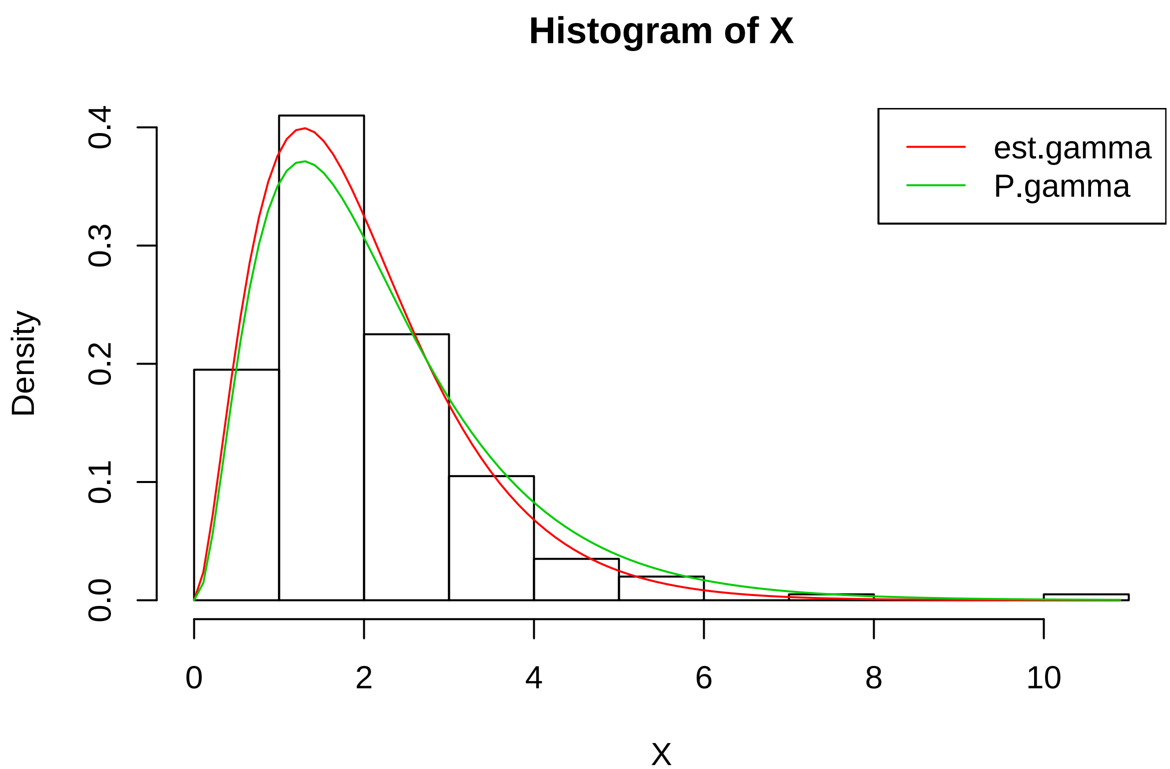




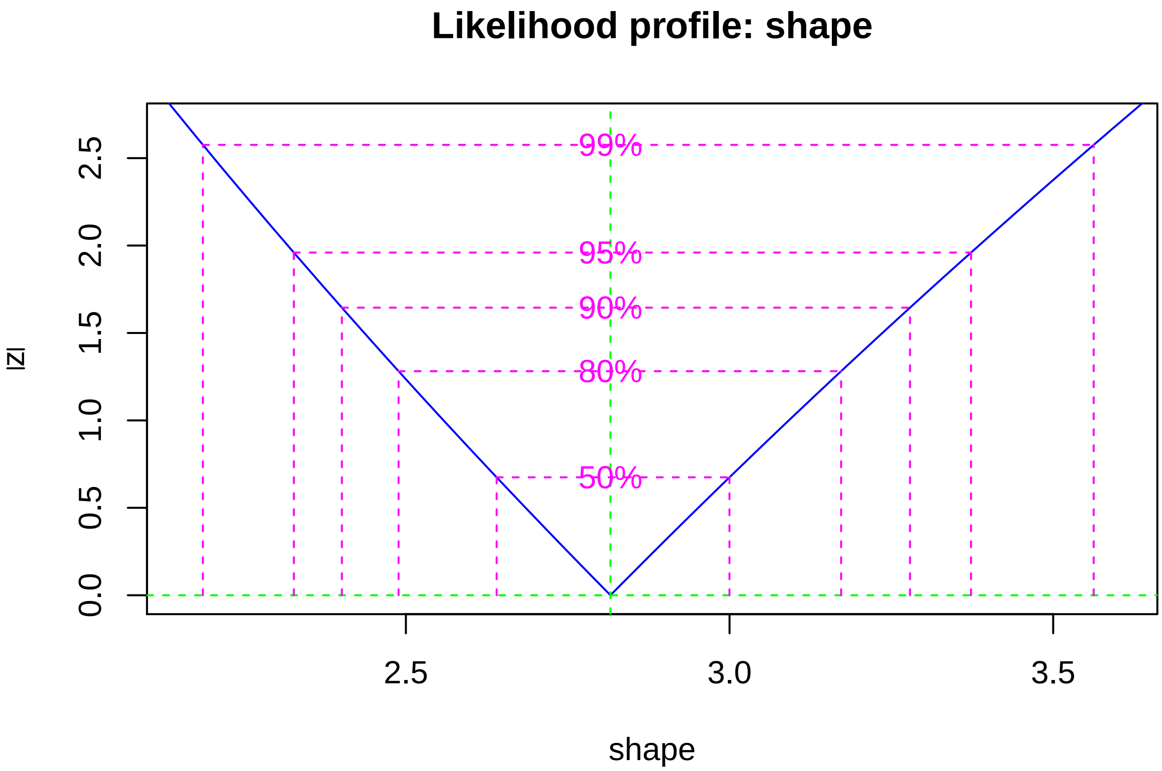


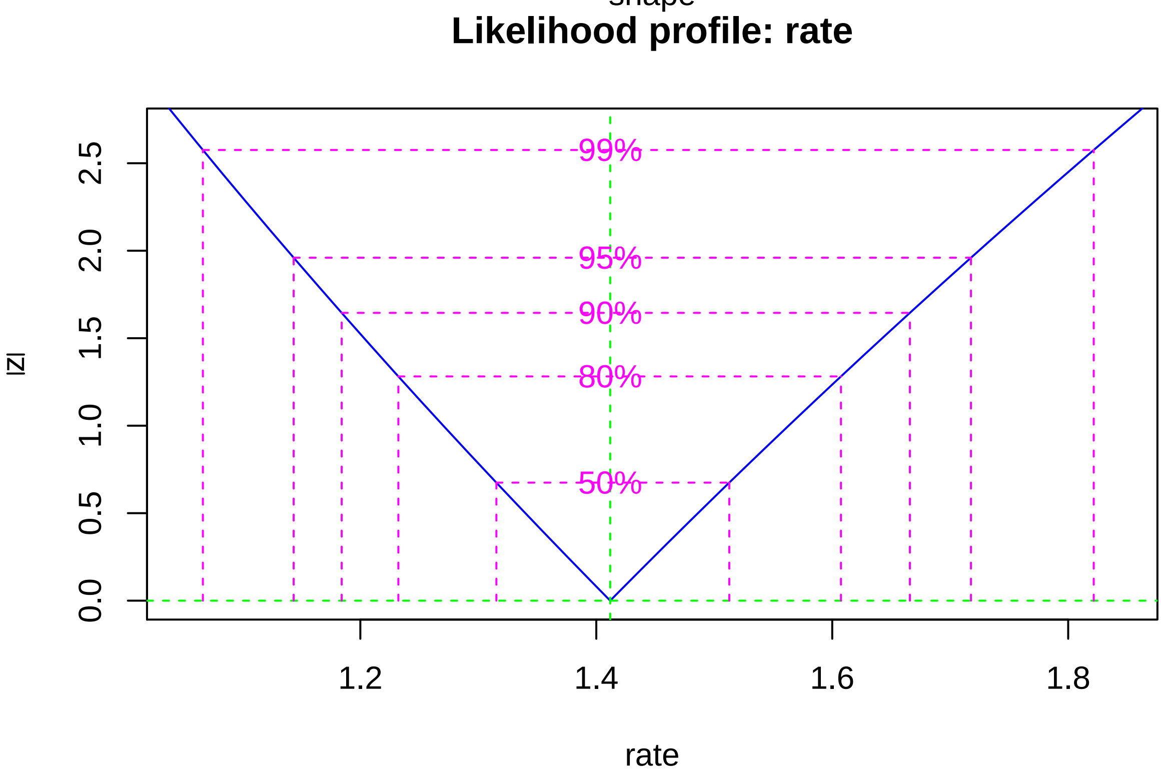
Затем промоделируем $\gamma_p(3,5,2)$, а поскольку гистограмма на глаз не отличается от гистограммы гамма распределения, оценим параметры модели гамма распределения.

```
## shape.shape   rate.rate
##    2.816053    1.411795
```



Histogram of X
est.gamma
P.gamma
Density
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
X



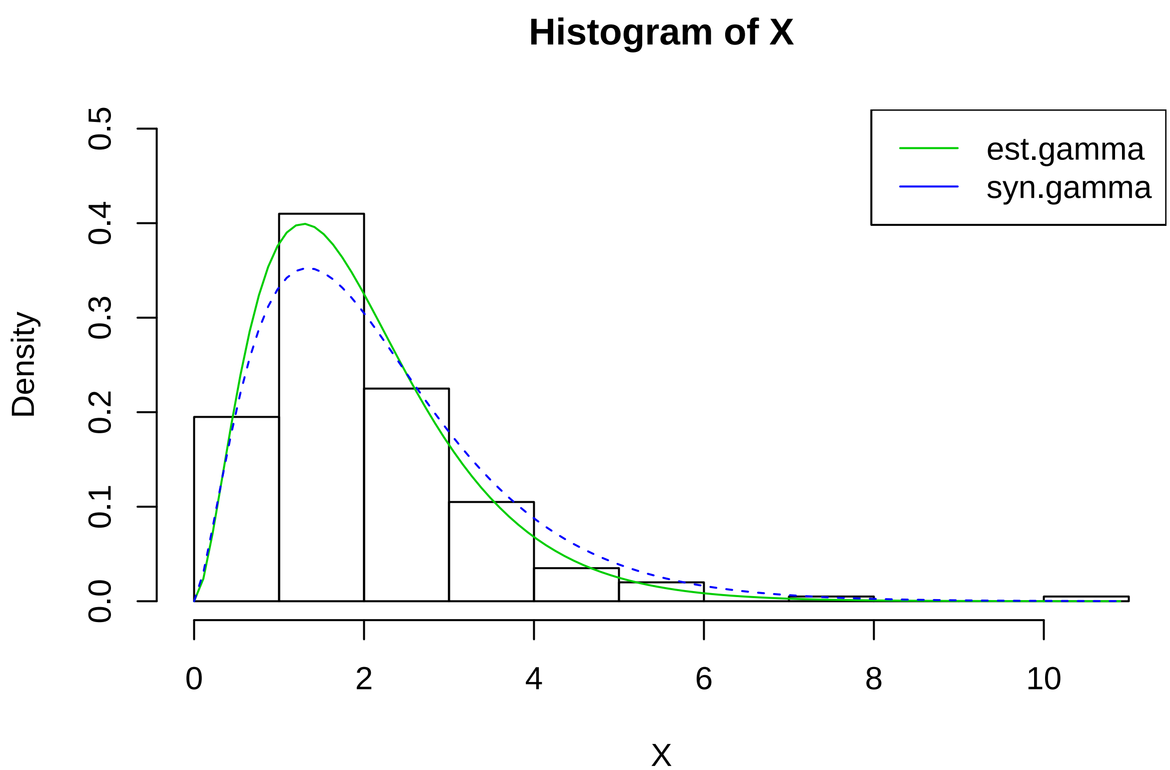


Теорема 1 Пусть имеются случайные величины X_1 и X_2, распределенные по $\gamma_p(\alpha_1, \lambda_1, \kappa_1)$ и $\gamma_p(\alpha_2, \lambda_2, \kappa_2)$, параметры α_2 и λ_2 для любого $\kappa_2 > 0$ выражаются через α_1, λ_1 и κ_1 следующим образом:

$$\lambda_2 = \left(\theta\left(\psi\left(\lambda_1 + \theta\right) - \psi(\lambda_1)\right)\right)^{-1}, \alpha_2 = \frac{\lambda_2 \alpha_1^{\theta} \Gamma(\lambda_1)}{\Gamma(\lambda_1 + \theta)}, \text{ где } \theta = \frac{\kappa_2}{\kappa_1}.$$

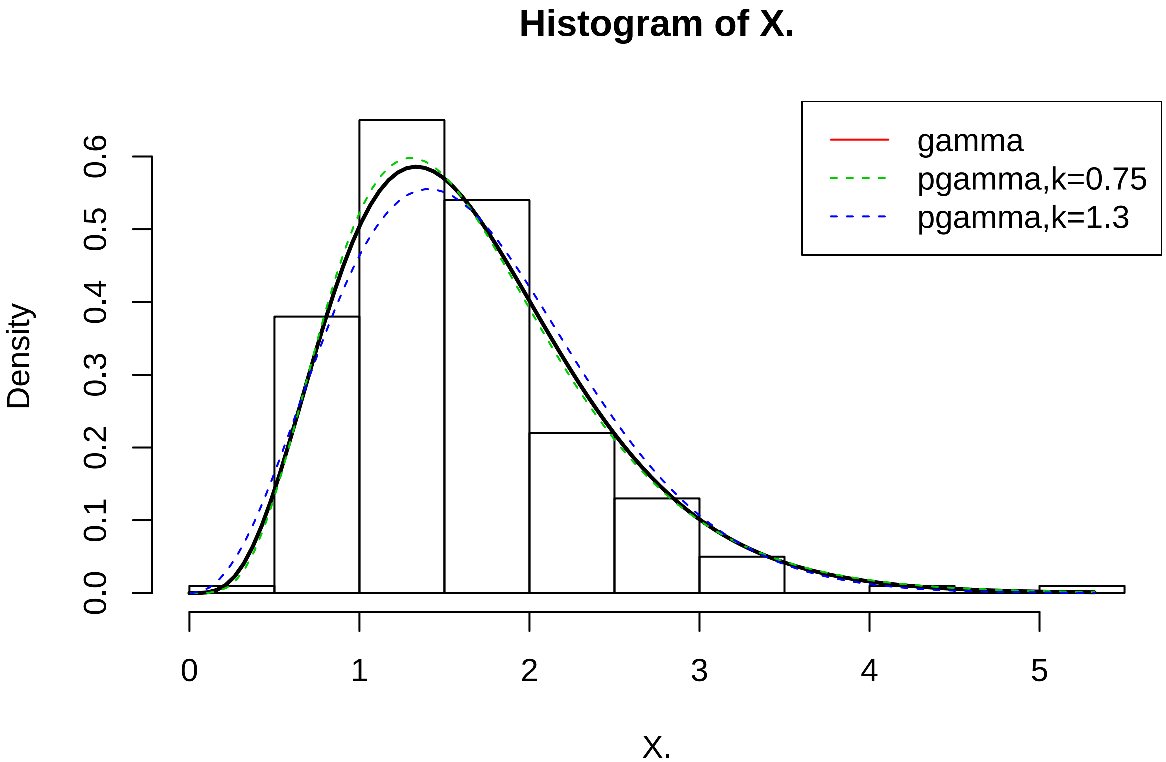
Тогда различие между плотностями минимально и $f_2(x) \sim f_1(x)$.

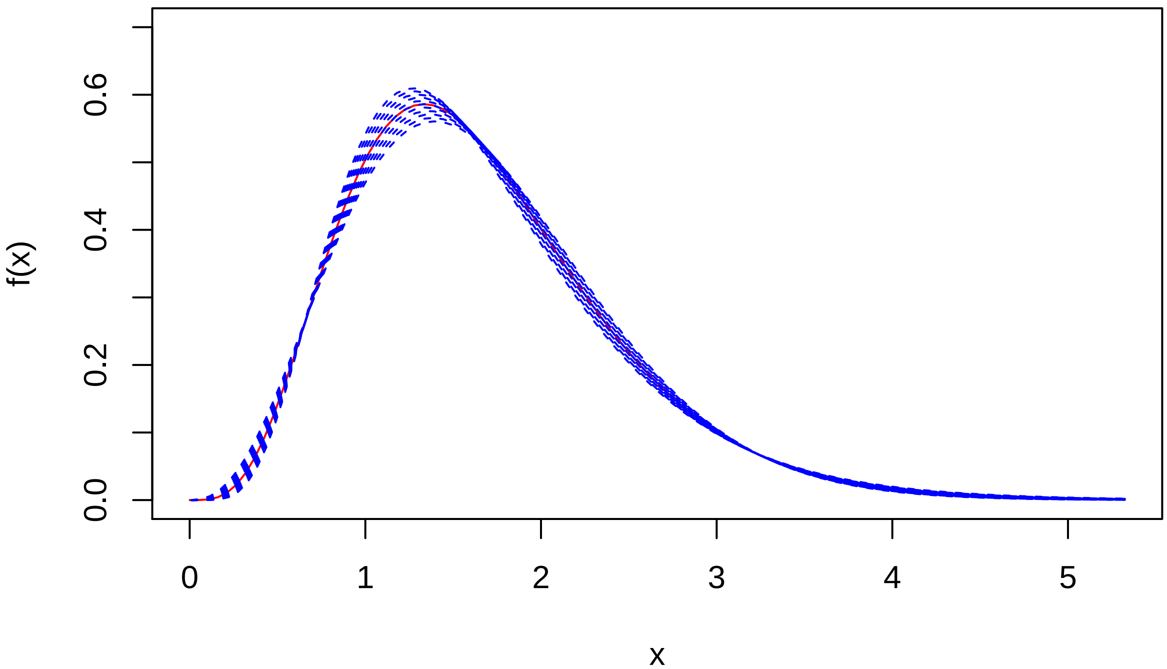
```
#синонимичный переход
Syn<-function(params,kappa2)
{

  lambda<-params[1];alpha<-params[2];kappa<-params[3];
theta<-kappa2/kappa

lam2<-1/(theta*(psigamma(lambda+theta, deriv = 0) -psigamma(lambda, deriv = 0) ))

alpha2<-lam2*alpha^theta*gamma(lambda)/gamma(lambda+theta)

c(lam2,alpha2,kappa2)
}

#----------------------
params<-c(lambda,alpha,kappa)
params2<-Syn(params,kappa2=1);

f2<-function(x)dgamma(x,shape=params.[1],rate=params.[2])
f2.<-function(x)dgamma(x,shape=params2[1],rate=params2[2])

hist(X,freq =FALSE,ylim=c(0,0.5))
curve(f2,0,max(X),col=3,add=TRUE)
curve(f2.,0,max(X),col=4,lty=2,add=TRUE)
legend('topright',c("est.gamma","syn.gamma"),col=c(3,4),lty=c(1,1))
```



Аналогично в обратную сторону. Если есть гамма распределение, то к нему можно подобрать степенное гамма практически с любым параметром.

```r
params2.<-Syn(c(lambda,alpha,1),0.75);
params3.<-Syn(c(lambda,alpha,1),1.6);

hist(X.,freq=FALSE)
f0<-function(x)dgamma(x,shape=params0[1],rate = params0[2])
curve(f0,0,max(X.),col=1,ylim=c(0,0.7),lwd=2,add=TRUE)
f..<-function(x)dpgamma(x,params2.)
curve(f..,0.01,max(X.),col=3,lty=2,add=TRUE)
f...<-function(x)dpgamma(x,params3.)
curve(f...,0.01,max(X.),col=4,lty=2,add=TRUE)
legend('topright',c("gamma","pgamma,k=0.75","pgamma,k=1.3"),col=c(2,3,4),lty=c(1,2,2))
```





Номинативное распределение

Для величины X_2, имеющей распределение $\gamma_p(\alpha_2, \lambda_2, \kappa_2)$ дифференциальная энтропия H_12 имеет вид:

$$H_{12} = -(\log\kappa_2 + \lambda_2\log\alpha_2 - \ln\Gamma(\lambda_2)) - (\kappa_2\lambda_2 - 1)\mathbf{E}\log X_1 + \alpha_2\mathbf{E}X_1^{\kappa_2},$$

где логарифмический и степенной моменты равны соответственно

$$\mathbf{E}\ln X_1 = \frac{1}{\kappa_1}\left(\psi(\lambda_1) - \ln\alpha_1\right), \quad \mathbf{E}X_1^{\kappa_2} = \frac{\Gamma(\lambda_1 + \frac{\kappa_2}{\kappa_1})}{\Gamma(\lambda_1)\alpha_1^{\frac{\kappa_2}{\kappa_1}}}.$$

Через $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ обозначена дигамма-функция или производная от логарифма гамма-функции.

```
ELN<-function(params2.)(psigamma(params2.[1])-log(params2.[2]))/params2.[3]

H12<-function(params1.,params2.){
  v1<- log(params2.[3]) +params2.[1]*log(params2.[2])-log(gamma(params2.[1])); v1
  v2<- (params2.[3]*params2.[1]-1)*ELN(params1.);v2
  theta<-params2.[3]/params1.[3]
  v3<- params2.[2]*gamma(params1.[1]+theta)/gamma(params1.[1])/params1.[2]^theta;v3
  -v1-v2+v3
}
```

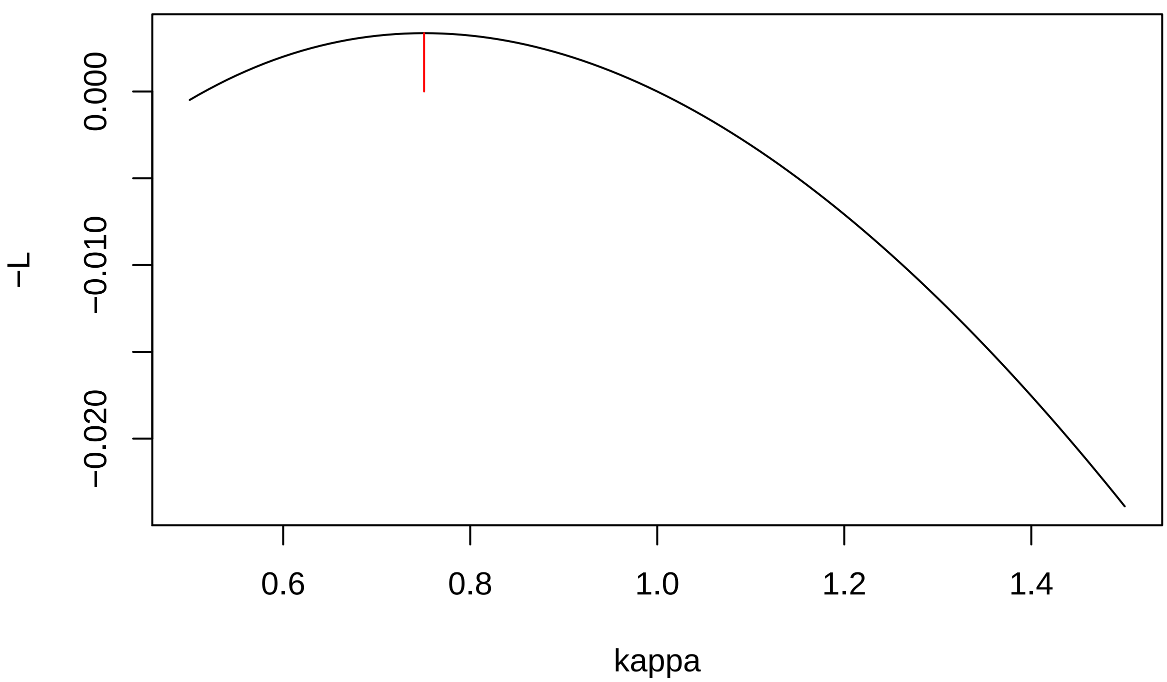
Метрики различия между распределениями

$$I_1 = (H_{12} - H_{11}) + (H_{21} - H_{22}),$$

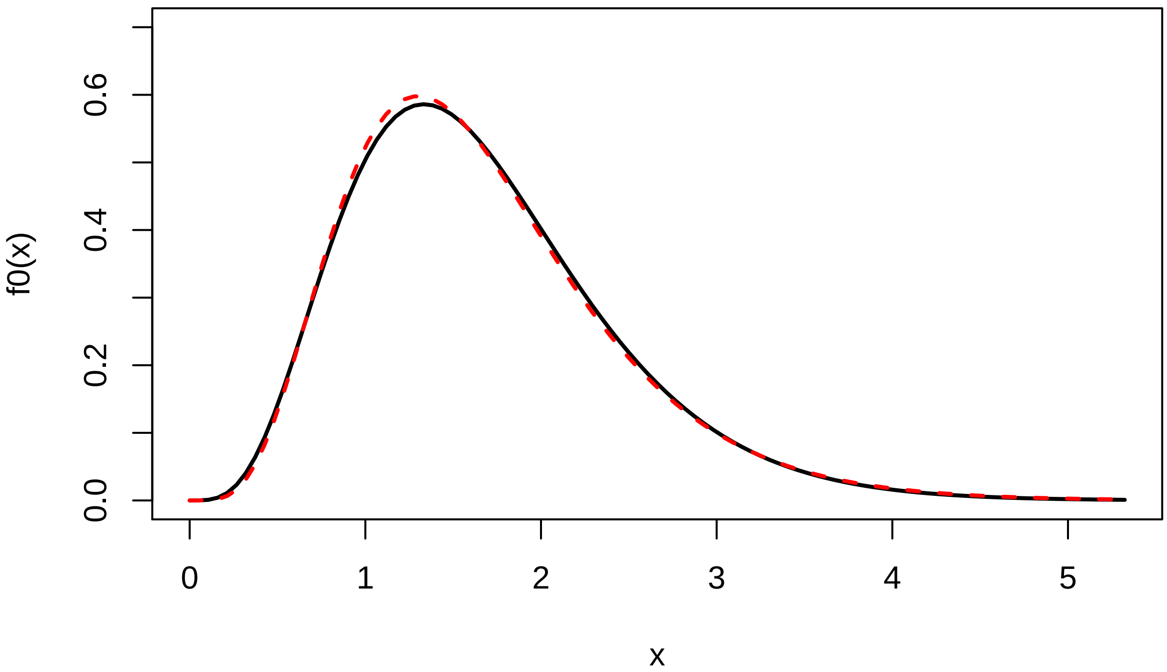
$$I_2 = (H_{11} - H_{22}) + (H_{12} - H_{21})$$

В качестве номинативного выберем распределение с максимальным значением I_2.

```
I2<-function(params1.,params2.)
  H12(params1.,params1.)-H12(params2.,params2.)+
  H12(params1.,params2.)-H12(params2.,params1.)


FK<-function(x) -I2(params1.,Syn( params1.,x) )


  params1.<-as.numeric(params0)

  l<-seq(from=0.5,to=1.5,by=0.01)

  L<-sapply(l,function(x) FK(x) )

  res<-optimise(FK,c(-0.5,2))

  Min.x<-res$minimum

 plot(l,-L,type="l",xlab="kappa")
 lines(c(Min.x,Min.x), c(0,-FK(Min.x)),type="l",col=2)
```




```
 f0<-function(x)dgamma(x,shape=params0[1],rate = params0[2])
 curve(f0,0,max(X.),col=1,ylim=c(0,0.7),lwd=2)
 f..<-function(x)dpgamma(x,Syn( params1.,Min.x))
 curve(f..,0,max(X.),col=2,ylim=c(0,0.7),lwd=2,lty=2,add=TRUE)
```


f0(x)
0.0
0.2
0.4
0.6
0
1
2
3
4
5
x