■ **以下の問題を考えてみてください**

コインを３回投げました。
表が２回、裏が１回出ました。
このコインの表が出る確率は？

事後分布

■　事象 y（再掲）

- １回目の**試行**（コイン投げのことですね）で・・・

「表が出た」という事象を $y_1 = 1$　　「裏が出た」という事象を $y_1 = 0$

と表しましょう。

- 確率変数

y_1は、0か1が出るということなので、確率変数 y_1 は・・・

$$y_1 \in \{0,1\}$$

２回目，3回目も

$$y_2 \in \{0,1\}, y_3 \in \{0,1\}$$

と表し、

$$y = (y1, y2, y3)$$

と表記しましょう。つまり，例題の場合（表、表、裏）であったら

$$y = (1,1,0)$$

ですね。

■ 事後分布 $\Pr(\boldsymbol{p}|y)$

- 尤度とは・・・

  $\Pr(y|\boldsymbol{p})$

  パラメータ $\boldsymbol{p}$ を前提とした場合に 確率変数 $\boldsymbol{y}$ が生じる確率。

- **事後分布(posterior distribution)**とは

  $\Pr(\boldsymbol{p}|y)$

  確率変数 $\boldsymbol{y}$ が観察された場合に，パラメータ$\boldsymbol{p}$ をがとりうる値の確立，可能性。

  $\boldsymbol{y}$ と $\boldsymbol{p}$ をひっくり返しただけど，意味が全然違う！

  尤度は、原因 $\boldsymbol{p}$ から結果 $\boldsymbol{y}$ が生じると考えられるから、定義が簡単。しかし，事後分布は、結果 $\boldsymbol{y}$ から原因 $\boldsymbol{p}$ をかき集めて定義しなければならない。

事後分布

■ 事後分布 $\Pr(p|y)$ ・・・続き

事後分布を求めてみよう・・・

その前に，**事前分布（prior distribution)**というのも知らないといけない。

- **事前分布** $\Pr(p)$

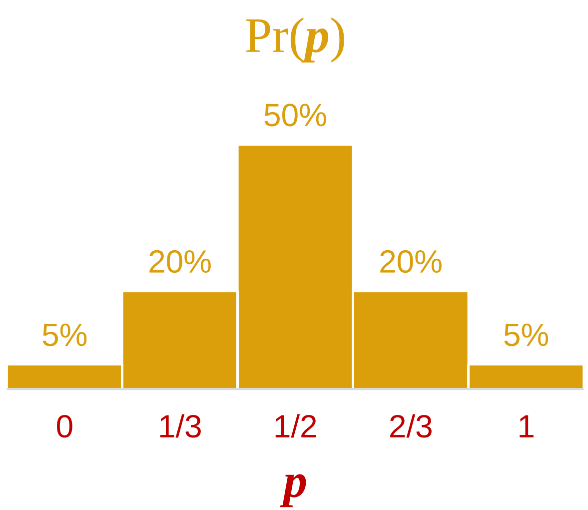
実際にどんな y が観察できたかはほっといて，そもそも p は，どんな割合で出現するのか？

- 例えば・・・

このコインをつくるおっちゃんの技術は確かで，50%はきっちり $p=1/2$ のコインをつくるが，たまには手元が狂うこともある。

話を簡単にするために $p=0, 1/3. 1/2 , 2/3, 1$の5つしかないとします。おっちゃんがつくるコインの p の割合、つまり、事前確率 $\Pr(p)$ は以下。

p	$\Pr(p)$
0	5%
1/3	20%
1/2	50%
2/3	20%
1	5%



■ 同時分布$\Pr(\boldsymbol{p}, y)$

事前確率 $\Pr(\boldsymbol{p})$ がわかれば、どんなpでどんなyが生じるか、その確率がわかる。

これを**同時分布（joint distribution）**といって、以下で求めることができる。

$$\Pr(\boldsymbol{p}, y) = \Pr(y|\boldsymbol{p})\ \mathbf{Pr}(\boldsymbol{p})$$

つまり、

同時分布＝尤度**×事前分布**

ということ。

■ 同時分布 $\Pr(p, y)$ ･･･続き

先ほど尤度の表から、同時分布を求めてみると････

事前分布 ＼ $y=$	(1,1,1)	(1,1,0)	･･･	(0,0,1)	(0,0,0)	計
$\Pr(p)$	$p^3\Pr(p)$	$p^2(1-p)\Pr(p)$	･･･	$p(1-p)^2\Pr(p)$	$(1-p)^3\Pr(p)$	$\Pr(p)$
$\Pr(p=0)=0.05$	0×0.05	0×0.05	･･･	0×0.05	1×0.05	0.05
$\Pr\left(p=\frac{1}{3}\right)=0.2$	$\frac{1}{27}$×0.2	$\frac{2}{27}$×0.2	･･･	$\frac{4}{27}$×0.2	$\frac{8}{27}$×0.2	0.2
$\Pr\left(p=\frac{1}{2}\right)=0.5$	$\frac{1}{8}$×0.5	$\frac{1}{8}$×0.5	･･･	$\frac{1}{8}$×0.5	$\frac{1}{8}$×0.5	0.5
$\Pr\left(p=\frac{2}{3}\right)=0.2$	$\frac{8}{27}$×0.2	$\frac{4}{27}$×0.2	･･･	$\frac{2}{27}$×0.2	$\frac{1}{27}$×0.2	0.2
$\Pr(p=1)=0.05$	1×0.05	0×0.05	･･･	0×0.05	0×0.05	0.05
$\Pr(y)$	$\Pr(y=(1,1,1))$	$\Pr(y=(1,1,0))$	･･･	$\Pr(y=(0,0,1))$	$\Pr(y=(0,0,0))$	1

■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

それでは、あらためて y=(1,1,0)の事後分布を求めてみよう。

y=(1,1,0) が生じたことがわかっていて、そのうちのどの $\boldsymbol{p}$ が生じたか、という確率なので，その事後確率は・・・

$$\Pr(\boldsymbol{p}|y) = \frac{\Pr(y, \boldsymbol{p})}{\Pr(y = (1,1,0))}$$

■ 事後分布 $\Pr(p|y)$ ・・・続き

このうち，分母の $\Pr(y=c(1,1,0))$ は， 以下の赤枠

事前分布 ＼ $y=$	(1,1,1)	(1,1,0)	・・・	(0,0,1)	(0,0,0)	計
$\Pr(p)$	$p^3\Pr(p)$	$p^2(1-p)\Pr(p)$	・・・	$p(1-p)^2\Pr(p)$	$(1-p)^3\Pr(p)$	$\Pr(p)$
$\Pr(p=0)$=0.05	0×0.05	0×0.05	・・・	0×0.05	1×0.05	0.05
$\Pr\left(p=\frac{1}{3}\right)$=0.2	$\frac{1}{27}$×0.2	$\frac{2}{27}$×0.2	・・・	$\frac{4}{27}$×0.2	$\frac{8}{27}$×0.2	0.2
$\Pr\left(p=\frac{1}{2}\right)$=0.5	$\frac{1}{8}$×0.5	$\frac{1}{8}$×0.5	・・・	$\frac{1}{8}$×0.5	$\frac{1}{8}$×0.5	0.5
$\Pr\left(p=\frac{2}{3}\right)$=0.2	$\frac{8}{27}$×0.2	$\frac{4}{27}$×0.2	・・・	$\frac{2}{27}$×0.2	$\frac{1}{27}$×0.2	0.2
$\Pr(p=1)$=0.05	1×0.05	0×0.05	・・・	0×0.05	0×0.05	0.05
$\Pr(y)$	$\Pr(y=(1,1,1))$	$\Pr(y=(1,1,0))$	・・・	$\Pr(y=(0,0,1))$	$\Pr(y=(0,0,0))$	1

■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

なので、y=(1,1,0) のが観察された場合の事後分布は,

$$\Pr(p=0|y=(1,1,0)) = \frac{0\times\mathbf{0.05}}{0\times\mathbf{0.05}+\frac{2}{27}\times\mathbf{0.2}+\frac{1}{8}\times\mathbf{0.5}+\frac{4}{27}\times\mathbf{0.2}+0\times\mathbf{0.05}} = 0$$

$$\Pr\left(p=\frac{1}{3}\middle|y=(1,1,0)\right) = \frac{\frac{2}{27}\times\mathbf{0.2}}{0\times\mathbf{0.05}+\frac{2}{27}\times\mathbf{0.2}+\frac{1}{8}\times\mathbf{0.5}+\frac{4}{27}\times\mathbf{0.2}+0\times\mathbf{0.05}} = 0.138528$$

$$\Pr\left(p=\frac{1}{2}\middle|y=(1,1,0)\right) = \frac{\frac{1}{8}\times\mathbf{0.5}}{0\times\mathbf{0.05}+\frac{2}{27}\times\mathbf{0.2}+\frac{1}{8}\times\mathbf{0.5}+\frac{4}{27}\times\mathbf{0.2}+0\times\mathbf{0.05}} = 0.584416$$

$$\Pr\left(p=\frac{2}{3}\middle|y=(1,1,0)\right) = \frac{\frac{4}{27}\times\mathbf{0.2}}{0\times\mathbf{0.05}+\frac{2}{27}\times\mathbf{0.2}+\frac{1}{8}\times\mathbf{0.5}+\frac{4}{27}\times\mathbf{0.2}+0\times\mathbf{0.05}} = 0.277056$$

$$\Pr(p=1|y=(1,1,0)) = \frac{0\times\mathbf{0.05}}{0\times\mathbf{0.05}+\frac{2}{27}\times\mathbf{0.2}+\frac{1}{8}\times\mathbf{0.5}+\frac{4}{27}\times\mathbf{0.2}+0\times\mathbf{0.05}} = 0$$

・・・ということ。

$\boldsymbol{p}=\frac{\mathbf{1}}{\mathbf{2}}$ が最頻値と考えられます。

■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

● もう少し一般的に説明すると・・・

事後分布は以下で定義されます。

$$\Pr(\boldsymbol{p}|y)=\frac{\Pr(y|\boldsymbol{p})\,\mathbf{Pr}(\boldsymbol{p})}{\Pr(y)}$$

ここで分母の $\Pr(y)$ は、$\boldsymbol{p}$ とは関わりなく一定なので， $\Pr(y)$を省略しても、$\Pr(\boldsymbol{p}|y)$の分布は同じようなものになる。もう少し、正確に言うと、 $\boldsymbol{p}$ による、$\Pr(\boldsymbol{p}|y)$の大小関係に影響を与えない。

なので、事後分布の実際の計算では、以下のように扱われる。

$$\Pr(\boldsymbol{p}|y)\propto\Pr(y|\boldsymbol{p})\,\mathbf{Pr}(\boldsymbol{p})$$

ここで、∝は、「比例する」という意味。言い換えると、

事後分布∝ 尤度×**事前分布**

■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

事前分布ももう少し一般的に、連続的な場合で考えましょう。

● 事前分布

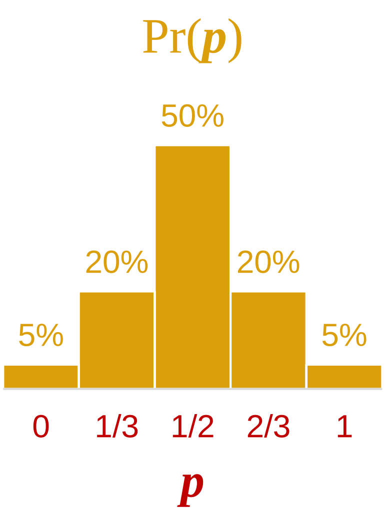
0.5が出やすく、0 や1は出にくいような、計算が簡単そうな事前分布を考えます。

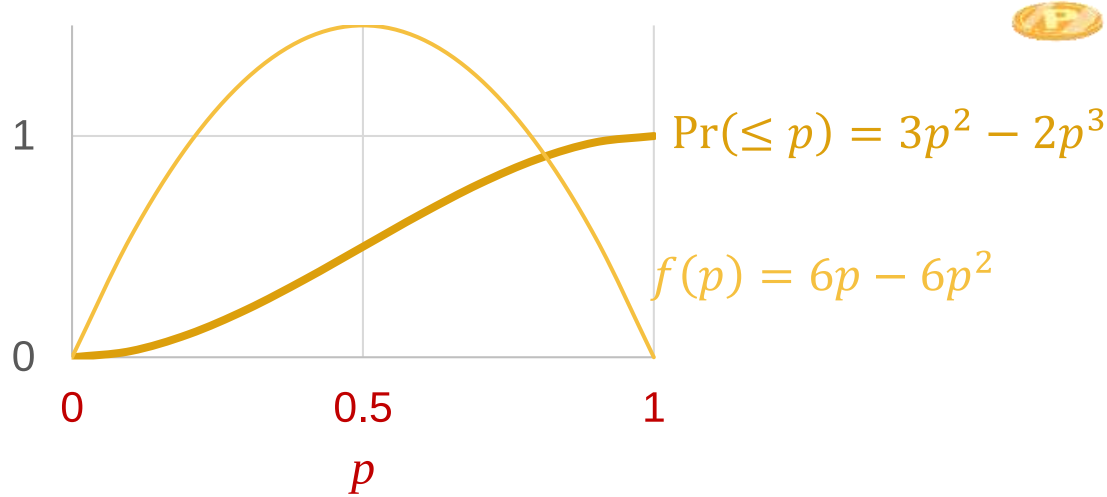
累積分布が・・・

$$\Pr(\leq p) = 3p^2 - 2p^3$$

密度関数は・・・

$$f(p) = 6p - 6p^2$$



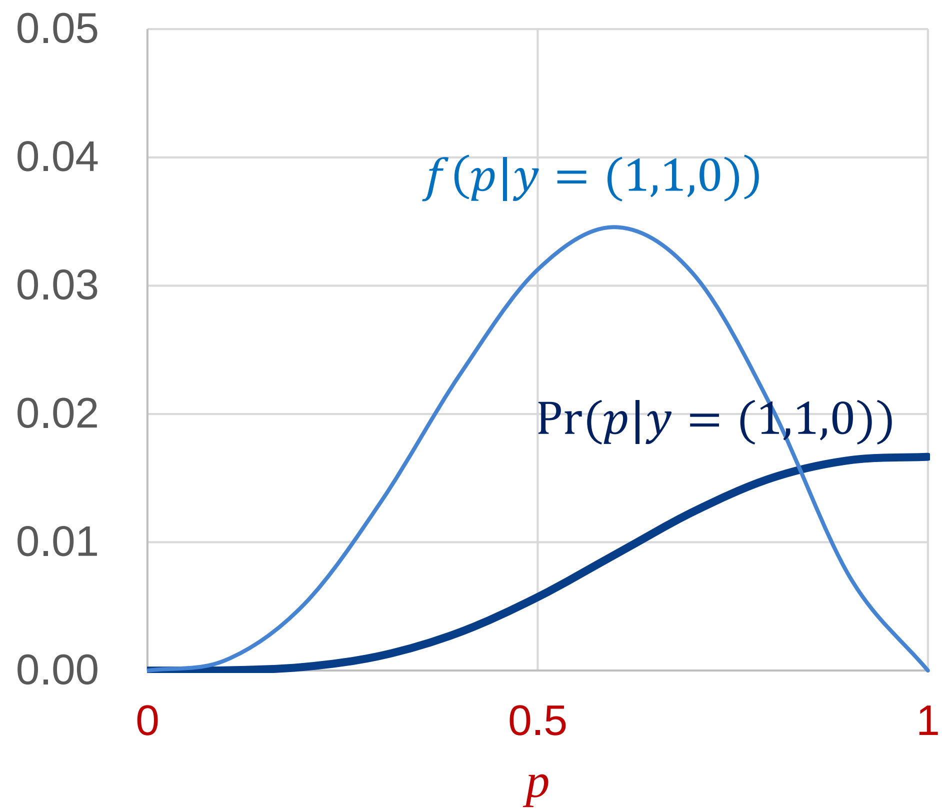


■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

● $y = (1,1,0)$が観察された場合

事後分布の確率密度は、以下のように扱われる。

$$
\begin{aligned}
& f(\boldsymbol{p}|y = (1,1,0)) \\
& = \frac{\Pr(y|\boldsymbol{p})\, f(p)}{\int_{p=0}^{1} \Pr(y|p)\, f(p) dp} \\
& \propto \Pr(y|\boldsymbol{p})\, f(p) \\
& = \boldsymbol{p}^2(1-\boldsymbol{p}) \cdot 6p(1-p) \\
& \propto \boldsymbol{p}^2(1-\boldsymbol{p}) \cdot p(1-p) \\
& = \boldsymbol{p}^3(1-\boldsymbol{p})^2
\end{aligned}
$$



比例で計算しているので、累積分布が
１になりませんが、形は同じです。

■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

ちなみに、分母をわざわざ計算してみると・・・

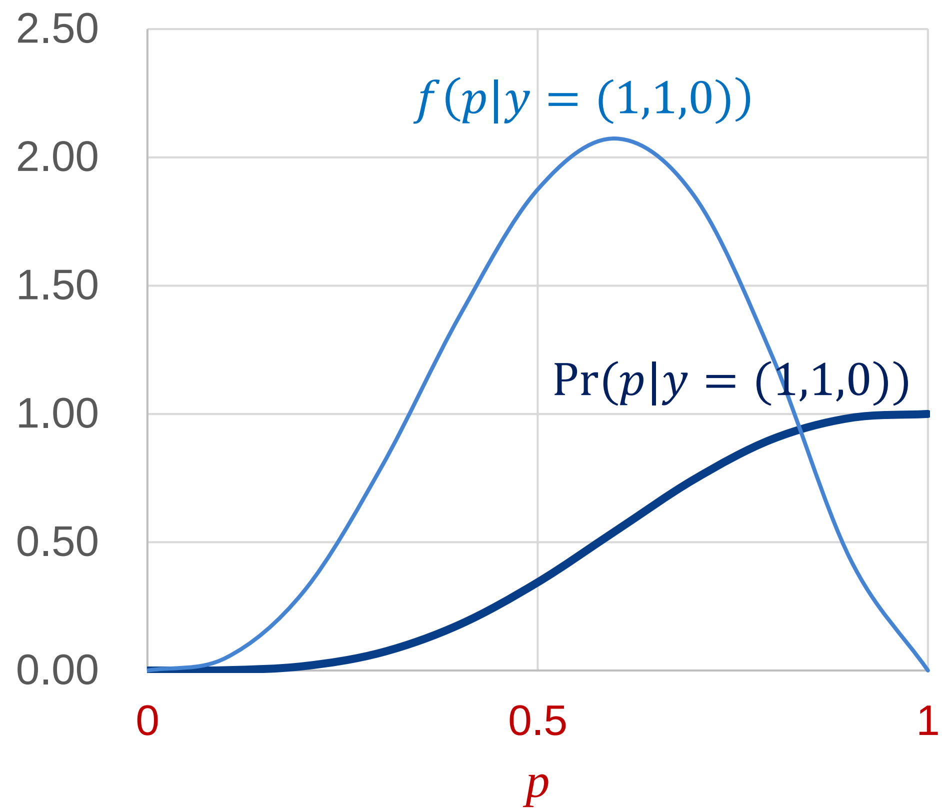
$$\int_{p=0}^{1} \Pr(y = (1,1,0)|p)\, f(p)\, dp = 0.1$$

でした。

また、事後分布の計算の途中で6を割っている

そこで、前のグラフの値を60倍してみると・・・

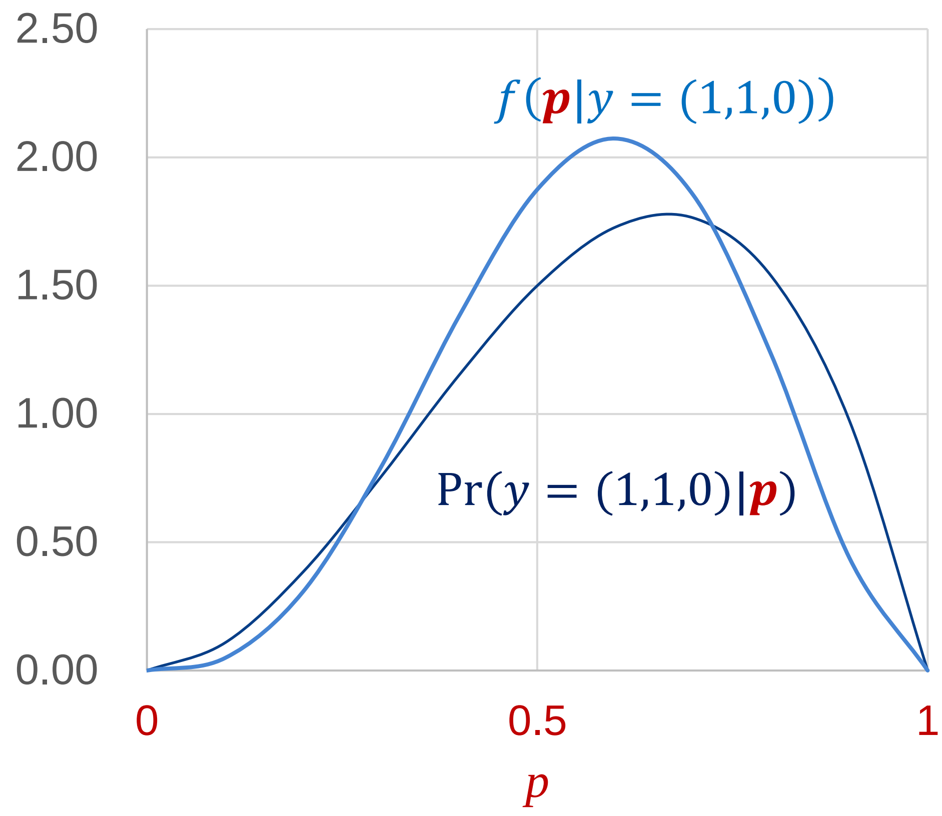
ちゃんと累積分布が最後１になっていますね。



■ 事後分布 $\Pr(\boldsymbol{p}|y)$ ・・・続き

尤度関数と比較してみると・・・

事後分布の方が少し0.5の方に寄っている。



■ 事前分布って何？

ところで、どっかから持ってきたのかわからない事前分布

$$\mathbf{Pr}(\boldsymbol{p})$$

これは何者？と思わなかったでしょうか？

じつは、ベイズ統計学に対する批判の1つでもあります。

勝手にへんな分布を仮定せずに、例えば、事前分布の密度関数に

$$\boldsymbol{f}(\boldsymbol{p}) = \mathbf{1}$$

という**一様分布**を用いると、これは事後分布は尤度だけになり、**最尤法と同じ**となります。

でも、使える事前情報は使わないともったいないよね、という考え方もあります。

例えば・・・

■ 事前分布って何？

- UFOにやられた？

  ある人が不幸なことに亡くなられました。

  原因は、UFOにやられた($\boldsymbol{p}$=1)か、風邪をひいたか($\boldsymbol{p}$=2)のどれかしかなく、平穏な時($\boldsymbol{p}$=3)に亡くなることはほぼないとしましょう。

- 確率変数

  無事 $y = 0$ 亡くなられた $y = 1$

- 尤度

  UFOにやられて亡くなる $\Pr(y = 1|p = 1) = 0.999$

  風邪をひいて亡くなる $\Pr(y = 1|p = 2) = 0.01$

  平穏なのに無くなる $\Pr(y = 1|p = 3) = 0.000001$

- 最尤法だと

  UFOにやられて亡くなった可能性が極めて高い！

■ 事前分布って何？・・・続き

● 事前分布

UFOにやられた　$\Pr(p=1)$= 1 00億分の1

風邪をひいた　$\Pr(p=2)$=100分の1

平穏　$\Pr(p=3)$=1－100億分の1－100分の1

● 事後分布

UFOにやられた　$\Pr(p=1|y=1) \propto 0.999$ ×100億分の 1 ≒0

風邪をひいた　$\Pr(p=2|y=1) \propto 0.01$ ×0.01=0.0001

平穏　$\Pr(p=2|y=1) \propto 0.000001$ ×(1-億分の1-0.01)=0.00000099

● 事後分布からは

風邪をひいて亡くなられたが最頻値

$y=$ / 事前分布	1	0
$\Pr(\mathbf{UFO}) = \dfrac{1}{100億}$	$0.999 \times \dfrac{1}{100億}$	$0.001 \times \dfrac{1}{100億}$
$\Pr(風邪) = \dfrac{1}{100}$	$0.01 \times \dfrac{1}{100}$	$0.99 \times \dfrac{1}{100}$
$\Pr(平穏) = 1 - \dfrac{1}{100億} - \dfrac{1}{100}$	$0.000001 \times \left(1 - \dfrac{1}{100億} - \dfrac{1}{100}\right)$	$0.999999 \times \left(1 - \dfrac{1}{100億} - \dfrac{1}{100}\right)$

■ 事前分布って何？・・・続き

- 事前分布が間違っていたら・・・

そりゃUFOのような極端な場合はいいけど、コイン投げのような微妙な場合は、事後分布は尤度を事前分布がひっぱっている。

その事前分布が間違っているの場合はどうなるのか？

- 試行回数を増やしてみる

コイン投げを30回に増やしてみたところ、表が20回、裏が10回だった。

- 事後分布は、

$$f(p|y) \propto p^{20}(1-p)^{10}\boldsymbol{p(1-p)}$$

で、試行回数が多くなるにつれ、事前分布の影響は薄れていく。